



PATENT ABSTRACTS OF JAPAN

(11) Publication number: 2000090111 A

(43) Date of publication of application: 31.03.00

(51) Int. Cl.

G06F 17/30

(21) Application number: 10260686

(22) Date of filing: 14.09.98

(71) Applicant: MATSUSHITA ELECTRIC IND CO LTD

(72) Inventor: SHIOMI RYUICHI

(54) INFORMATION RETRIEVAL AGENT DEVICE, AND COMPUTER- READABLE RECORDING MEDIUM RECORDED WITH PROGRAM EXHIBITING FUNCTION OF INFORMATION RETRIEVAL AGENT DEVICE

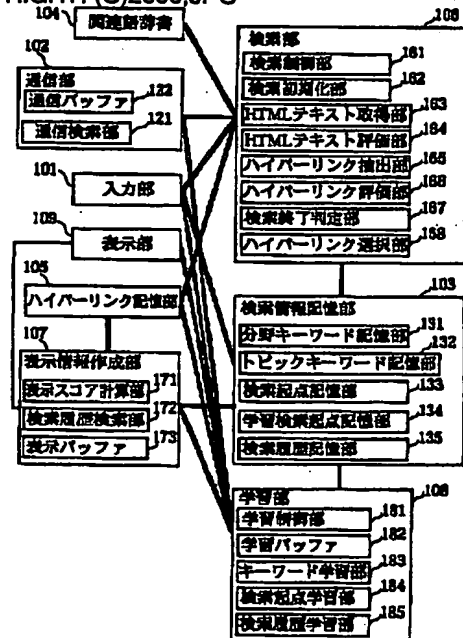
display information generation part 107 rearranges character strings stored in a hyperlink storage part 105 according to specific selection standards by referring to a retrieval history and displays their list at a display part 109.

(57) Abstract:

PROBLEM TO BE SOLVED: To gather more necessary WWW(worldwide web) pages within a given time and a storage area.

SOLUTION: A retrieval part 106 when receiving an instruction for retrieval from an input part 101 acquires the URL(uniform resource locator) of a retrieval starting point from a retrieval information storage part 103, acquires the corresponding HTML (hypertext markup language) text through a communication part 102 and analyzes it, and stores link destination URLs, anchor character strings, etc., in a hyperlink storage part 105. A character string is compared with key words stored in the retrieval information storage part 103 and an evaluated value of the degree of matching data specified by a link destination URL with the purpose of retrieval is calculated. The link destination URL having the largest evaluated value is extracted while a link destination URL having 'link', etc., in the character string, etc. The HTML text of the extracted URL is obtained and the retrieval is repeated up to a specific end condition. A

COPYRIGHT: (C)2000,JPO



(19)日本国特許庁(JP)

(12) 公開特許公報(A)

(11)特許出願公開番号

特開2000-90111

(P2000-90111A)

(43)公開日 平成12年3月31日(2000.3.31)

(51)IntCl.
G 0 6 F 17/30

識別記号

F I

G 0 6 F 15/40

ターム(参考)

3 8 0 Z 5 B 0 7 5

3 1 0 C

15/403

3 4 0 A

3 8 0 E

15/419

3 2 0

審査請求 未請求 請求項の数13 O L (全 33 頁)

(21)出願番号 特願平10-260686

(22)出願日 平成10年9月14日(1998.9.14)

(71)出願人 000005821

松下電器産業株式会社

大阪府門真市大字門真1006番地

(72)発明者 植見 隆一

大阪府門真市大字門真1006番地 松下電器
産業株式会社内

(74)代理人 100090446

弁理士 中島 司朗 (外1名)

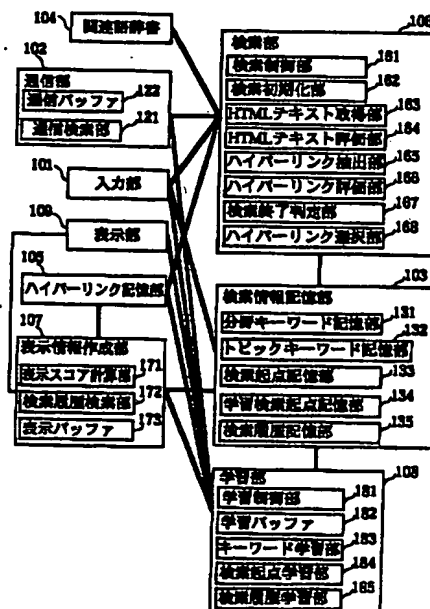
Fターム(参考) 5B075 KK02 KK37 ND03 NK02 NK04
NK44 NK54 PQ46 UU40

(54)【発明の名称】 情報検索エージェント装置及び情報検索エージェント装置の機能を発揮するプログラムを記録したコンピュータ読み取り可能な記録媒体

(57)【要約】

【課題】 与えられた時間や記憶領域の中でより必要なWWWページを収集する。

【解決手段】 検索部106は、入力部101から検索の指示を受けると、検索起点のURLを検索情報記憶部103から取得し、対応するHTMLテキストを通信部102を介して取得し、解析し、リンク先URLやアンカー文字列等をハイパーリンク記憶部105に記憶させる。検索情報記憶部103に記憶されているキーワードと文字列等とを比較し、リンク先URLで指定されるデータの検索目的合致度の評価値を計算する。文字列等に「リンク」等を含むリンク先URLを優先して、最高の評価値のリンク先URLを抽出する。抽出したURLのHTMLテキストを取得し、所定の終了条件まで検索を繰り返す。表示情報作成部107は、検索履歴を参照して、ハイパーリンク記憶部105に記憶されている文字列を所定の選択基準で並べ換え、表示部109に一覧表示させる。



【特許請求の範囲】

【請求項1】 ネットワーク及びシステム上のアドレスで指定されるデータを検索する情報検索エージェント装置であって、

アドレスと検索開始の指示とを受けるか又はアドレスの通知を受けると、当該アドレスで指定されるデータを取得するデータ取得手段と、

取得された複数のデータ要素からなるデータを記憶するデータ記憶手段と、

前記データ記憶手段に記憶されているデータを解析し、データ要素に含まれる他のデータへのリンク先アドレスと当該アドレスに関連する文字列とを抽出する抽出手段と、

抽出された文字列に基づいてリンク先アドレスで指定されるデータの選択適性の評価値を計算する計算手段と、

評価テーブルに、文字列と、計算された評価値とをリンク先アドレスごとに記載する評価テーブル記載手段と、前記評価テーブルに記載された評価値が最高のリンク先アドレスを前記データ取得手段に通知するデータ取得制御手段と、

前記データ取得制御手段の処理を前記評価テーブルの状況に応じて禁止する禁止手段と、

前記評価テーブルに記載された文字列を選択する選択手段と、

前記選択手段で選択された文字列を表示画面に一覧表示させる一覧表示制御手段とを備えることを特徴とする情報検索エージェント装置。

【請求項2】 請求項1記載の情報検索エージェント装置は、更に、

前記表示画面に表示された文字列の選択をユーザから受け取ると、対応するリンク先アドレスを前記データ取得手段に通知するデータ取得通知手段と、

前記データ取得通知手段によるアドレスの通知の後、前記データ記憶手段にデータが記憶されると、前記抽出手段の処理を禁止して、記憶されているデータを解析し、表示画面に表示させるデータ表示制御手段とを備えることを特徴とする情報検索エージェント装置。

【請求項3】 請求項1記載の情報検索エージェント装置は、更に、

ユーザからキーワードの入力を受け付けて記憶しているキーワード記憶手段と、

見出し語と、その見出し語に関連する関連語と、その関連語の見出し語に対する関連度とが登録された関連語辞書とを備え、

前記文字列は、リンク先アドレスで指定されるデータの内容を示す文字列であり、文字列には説明文が付されている場合があり、

前記計算手段は、

前記文字列及び説明文に含まれる単語が前記キーワード記憶手段に記憶されているキーワードと一致するとき、

評価値を大きくするキーワード評価部と、

前記文字列及び説明文に含まれる単語が前記キーワード記憶手段に記憶されているキーワードを前記関連語辞書の見出し語とする関連語と一致するとき、その関連度に応じて評価値を大きくする関連語評価部とを有することを特徴とする情報検索エージェント装置。

【請求項4】 請求項3記載の情報検索エージェント装置は、更に、

前記表示画面に表示された文字列の選択をユーザから受け付けると、当該文字列を形態素解析し、所定の条件により、当該文字列に含まれる単語をトピックキーワードに決定するトピックキーワード決定手段と、

前記決定されたトピックキーワードを記憶するトピックキーワード記憶手段とを備え、

前記計算手段は、

前記文字列及び説明文に含まれる単語が前記トピックキーワード記憶手段に記憶されているトピックキーワードと一致するとき、評価値を大きくするトピックキーワード評価部を有することを特徴とする情報検索エージェント装置。

【請求項5】 請求項1記載の情報検索エージェント装置の前記データ取得制御手段は、

前記評価テーブルに記載されたリンク先アドレスと文字列とのいずれかに「リンク」又は「link」の文字列が含まれているか否かを判断する判断部と、

前記判断部で含まれていると判断されたとき、当該リンク先アドレスを評価値が最高のリンク先アドレスに優先して前記データ取得手段に通知する優先通知部とを有することを特徴とする情報検索エージェント装置。

【請求項6】 請求項1記載の情報検索エージェント装置は、更に、

前記一覧表示制御手段によって文字列が表示されたとき、文字列ごとに表示回数を計数する計数手段と、

前記計数手段で計数された表示回数を対応するリンク先アドレスごとに記憶する検索履歴記憶手段とを備え、

前記選択手段は、

前記検索履歴記憶手段に記憶されている表示回数の多いリンク先アドレスに対応する文字列を劣後する劣後部を有することを特徴とする情報検索エージェント装置。

【請求項7】 請求項2記載の情報検索エージェント装置は、更に、

前記データ表示制御手段でデータが表示画面に表示されているときに、ユーザから興味の有無に関するボタン操作を受け付ける受付手段と、

前記受付手段で興味「有」を受け付けたとき選択されている文字列に対応するリンク先アドレスを検索起点として登録する検索起点登録手段と、

前記受付手段で興味「無」を受け付けたとき選択されている文字列に対応するリンク先アドレスが検索起点登録手段で登録されているとき、その検索起点であるリンク

先アドレスを削除する検索起点削除手段とを備え、前記データ取得手段は、検索開始の指示を受けたとき、検索起点登録手段に登録されている当該アドレスで指定されるデータを取得することを特徴とする情報検索エージェント装置。

【請求項8】 請求項4記載の情報検索エージェント装置は、更に、

前記トピックキーワード記憶手段に記憶されているトピックキーワードには、重要度が付されており、前記データ表示制御手段でデータが表示画面に表示されているときに、ユーザから興味の有無に関するボタン操作を受け付ける受付手段を備え、

前記トピックキーワード決定手段は、前記受付手段で興味「有」を受け付けたとき、対応する文字列から決定されたトピックキーワードの重要度を高くし、興味「無」を受け付けたとき、対応する文字列から決定されたトピックキーワードの重要度を低くする重要度評価部を有し、

前記トピックキーワード評価部は、この重要度を考慮することを特徴とする情報検索エージェント装置。

【請求項9】 請求項4記載の情報検索エージェント装置は、更に、

データ内容の分野別に名前を付し、検索開始時のアドレスと、キーワードと、トピックキーワードと、前記一覧表示制御手段での表示履歴とを含む検索情報を記憶する検索情報記憶手段と、

検索終了時に前記検索情報を更新する検索情報更新手段とを備え、

前記一覧表示制御手段は、ユーザから検索開始の指示を受けると、前記名前を表示画面に一覧表示させる初期表示部を有し、

前記データ取得制御手段は、

前記名前の選択を受けると、対応する検索開始時のアドレスを前記データ取得手段に通知する初期通知部を有することを特徴とする情報検索エージェント装置。

【請求項10】 請求項1記載の情報検索エージェント装置の前記計算手段は、

前記データ取得手段がデータを取得するのに要した時間を計時し、取得したデータ量とから送信速度を計算する送信速度計算部と、

前記送信速度計算部で計算された送信速度を同一サイトのアドレスごとに記憶する送信速度記憶部と、送信速度の速いサイトの評価値を大きくする送信速度評価部とを有することを特徴とする情報検索エージェント装置。

【請求項11】 請求項2記載の情報検索エージェント装置の前記計算手段は、

前記データ表示制御手段でデータが表示画面に表示されたとき、その表示回数を計数するデータ表示回数計数部と、

計数された表示回数を同一サイトのアドレス毎に記憶するデータ表示回数記憶部と、前記データ表示回数の多いサイトの評価値を大きくする表示回数評価部とを有することを特徴とする情報検索エージェント装置。

【請求項12】 ネットワークやシステム上の所定のアドレスで指定されるデータを検索する情報検索エージェント装置であって、

アドレスと検索開始の指示とを又はアドレスの通知を受けると、当該アドレスで指定されるデータを取得するデータ取得手段と、

取得された複数のデータ要素からなるデータを記憶するデータ記憶手段と、

前記データ記憶手段に記憶されているデータを解析し、データ要素に含まれる他のデータへのリンク先アドレスと当該アドレスで指定されるデータに関連する文字列とを抽出する抽出手段と、

抽出された文字列に基づいてリンク先アドレスで指定されるデータの選択適性の評価値を計算する計算手段と、

20 評価テーブルにリンク先アドレスと、文字列と、計算された評価値とをリンク先アドレスごとに記載する評価テーブル記載手段と、

前記評価テーブルに記載された評価値が最高のリンク先アドレスを前記データ取得手段に通知するデータ取得制御手段と、

前記データ取得制御手段の処理を前記評価テーブルの状況に応じて禁止する禁止手段と、

前記評価テーブルに記載された所定の文字列を選択する選択手段と、

30 前記選択手段で選択された文字列に対応するリンク先アドレスで指定されるデータを順次取得する一括データ取得手段と、

前記一括データ取得手段で取得されたデータを記憶する一括データ記憶手段とを備えることを特徴とする情報検索エージェント装置。

【請求項13】 コンピュータ読み取り可能な記録媒体であって、

ネットワークやシステム上の所定のアドレスで指定されるデータを検索する情報検索エージェント装置の、

40 アドレスと検索開始の指示とを又はアドレスの通知を受けると、当該アドレスで指定されるデータを取得するデータ取得手段と、

取得された複数のデータ要素からなるデータを記憶領域に書き込むデータ書込手段と、

前記データ記憶領域に記憶されているデータを解析し、データ要素に含まれる他のデータへのリンク先アドレスと当該アドレスで指定されるデータに関連する文字列とを抽出する抽出手段と、

50 抽出された文字列に基づいてリンク先アドレスで指定されるデータの選択適性の評価値を計算する計算手段と、

評価テーブルにリンク先アドレスと、文字列と、計算された評価値とをリンク先アドレスごとに記載する評価テーブル記載手段と、

前記評価テーブルに記載された評価値が最高のリンク先アドレスを前記データ取得手段に通知するデータ取得制御手段と、

前記データ取得制御手段の処理を前記評価テーブルの状況に応じて禁止する禁止手段と、

前記評価テーブルに記載された所定の文字列を選択する選択手段と、

前記選択手段で選択された文字列を表示画面に一覧表示させる一覧表示制御手段との各手段の機能を発揮させるプログラムを記録したコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、他のドキュメントへのリンクやハイパーリンクを定義するHTML (Hypertext Markup Language) テキストやXML (Extensible Markup Language) テキストを検索する情報検索エージェント装置に関する。

【0002】

【従来の技術】WWW (World Wide Web) のページを閲覧するビューワーとして、マイクロソフト社の「インターネットエクスプローラ」やネットスケープ社の「ネットスケープナビゲーター」などがある。これらのビューワーではURL (Uniform Resource Locaters) を入力したり、WWWページ中のハイパーリンクを指定することによって必要なWWWページを表示させることができる。

【0003】「インターネットエクスプローラ」や「ネットスケープナビゲーター」を用いて、必要なWWWページを検索する場合、WWWページを取得する時間、ユーザは待たなければならず、必要な情報をインタラクティブな操作で得るためには長い時間が必要である。

【0004】そのため、AIソフト社の「Web Whacker」などのオートパイロットソフトを用いると、URLとWWWページの取得範囲を指定すると、指定されたURLを起点としてハイパーリンクで指定されたWWWページをまとめてダウンロードすることができる。

【0005】これによって、オートパイロットソフトではまとめてWWWページをダウンロードしてくれるので、ユーザは長時間のインタラクティブな操作を回避できる。しかし、余分なWWWページまでダウンロードしてしまうため、多大の記憶領域が必要である。また、そのダウンロードに長い時間が必要である。また、ダウンロードしたWWWページから必要なものを取り出す作業も必要となる。

【0006】そこで、できるだけ重要なWWWページだけを収集する方法が、情報処理学会研究会報告98-NL-124-12「分野を特定した自動収集によるWWW情報検索」に紹介されている。

【0007】図52は、この文献に記述されている重要なWWWページを収集する動作を表すフローチャートである。

【0008】最初に、未取得のWWWページのURLを未取得URL集合Gとして、この集合Gに検索起点となるURLを登録する。また、このURLの評価値を高く設定しておく (S5201)。次に、取得済みのWWWページのURLを取得済みURL集合Kとして、これを空集合に初期化する (S5202)。集合Gから、最高評価値のURLを1つ取り出す (S5203)。取り出したURLの評価値が一定の値より小さければ検索を終了する (S5204)。

【0009】取り出したURLの評価値が一定の値より大きければ、取り出したURLに対応するHTMLテキストをネットワークを通じて取得する (S5205)。取り出したURLを集合Gから削除し (S5206)、集合Kに登録する (S5207)。取得したHTMLテキストを解析して、ハイパーリンク定義されているURLを抽出する (S5208)。抽出したURLに対してリンク重要度を式 (1) を用いて計算する (S5209)。

【0010】

【数1】

$$\text{リンク重要度} = C \cdot \frac{W}{S} \log_2 \frac{S}{(t+1)} \quad \dots (1)$$

ここで、wは、検索する分野を特定するためのキーワードを予め用意しておき、そのキーワードがHTMLテキスト中に出現する回数である。sはHTMLテキストの大きさ (バイト数) である。tは、HTMLテキスト中にハイパーリンク定義されているURLの数である。Cは適当な定数で、100が使用される。

【0011】次に、抽出したURLから1つを取り出し (S5210)、URLが集合Kに含まれる場合は、なにもしない (S5211)。URLが集合Kに含まれない場合は、URLが集合Gに含まれるか判定する (S5212)。含まれない場合、そのURLを集合Gに登録し、S5209で算出したリンク重要度をそのURLの評価値とする (S5213)。含まれる場合は、集合Gに既に含まれているURLの評価値にS5209で算出したリンク重要度を加える。これを抽出した全URLに対して行う (S5210～S5215)。抽出したURLへの動作が終了すると、制御をS5203に移し、さらに動作を継続する。

【0012】この手法では、検索する分野のより多くのWWWページからリンクされるURLのHTMLテキストが、より必要なページであることを前提として検索を

行うようなくみとなっている。

【0013】

【発明が解決しようとする課題】ところが、上記文献に示される重要なWWWページだけを収集する方法では、リンク重要度の式(1)からわかる通り、1つのHTMLテキストにハイパーリンク定義されているすべてのURLには同じリンク重要度が適用される。よって、検索初期段階では、すべてのURLを網羅的に検索し、ある程度のHTMLテキストが収集された状態でないと、真に重要度の高いWWWページが収集されるとは限らず、検索したい分野から無関係のURLが検索されてしまう可能性が高い。実際、上記文献では、5000個、30000個のWWWページを収集して、その中から必要なページを取り出さなければならない。

【0014】本発明は、上記課題に鑑み、ユーザが望むWWWページを少ない記憶領域を用いて、短時間に収集でき、かつ、通信コストを節約できる情報検索エージェント装置を提供することを目的とする。

【0015】更に、本発明の目的は、検索によって得られたWWWページから更に検索を続行する場合に、同一のWWWページを再検索する無駄を防止する情報検索エージェント装置を提供することである。

【0016】

【課題を解決するための手段】上記課題を解決するため本発明は、ネットワーク及びシステム上のアドレスで指定されるデータを検索する情報検索エージェント装置であって、アドレスと検索開始の指示とを受けると、当該アドレスの通知を受けると、当該アドレスで指定されるデータを取得するデータ取得手段と、取得された複数のデータ要素からなるデータを記憶するデータ記憶手段と、前記データ記憶手段に記憶されているデータを解析し、データ要素に含まれる他のデータへのリンク先アドレスと当該アドレスに関連する文字列とを抽出する抽出手段と、抽出された文字列に基づいてリンク先アドレスで指定されるデータの選択適性の評価値を計算する計算手段と、評価テーブルに、文字列と、計算された評価値とをリンク先アドレスごとに記載する評価テーブル記載手段と、前記評価テーブルに記載された評価値が最高のリンク先アドレスを前記データ取得手段に通知するデータ取得制御手段と、前記データ取得制御手段の処理を前記評価テーブルの状況に応じて禁止する禁止手段と、前記評価テーブルに記載された文字列を選択する選択手段と、前記選択手段で選択された文字列を表示画面に一覧表示させる一覧表示制御手段とを備えることとしている。

【0017】

【発明の実施の形態】以下、本発明に係る情報検索エージェント装置の実施の形態を、図面を参照しながら説明する。

(実施の形態1)

1. 1 構成

図1は、本発明に係る情報検索エージェント装置の実施の形態1の構成図である。

【0018】この情報検索エージェント装置は、入力部101と、通信部102と、検索情報記憶部103と、関連語辞書104と、ハイパーリンク記憶部105と、検索部106と、表示情報作成部107と、学習部108と、表示部109とを備えている。

1. 1. 1 入力部101

入力部101は、キーボードやマウスなどによって構成され、ユーザの指示を受け付ける。検索キーワードや検索起点のURLを受け付けると、検索情報記憶部103に記憶させる。

1. 1. 2 通信部102

通信部102は、通信検索部121と通信バッファ122とからなり、ネットワークとの通信を行う。

【0019】通信検索部121は、検索部106又は学習部108から与えられたURLでネットワークを介してHTMLテキストやXMLテキスト、イメージデータなどのリソースデータを取得する。

【0020】通信バッファ122は、磁気ディスクやメモリなどによって構成され、通信検索部121で取得されたリソースデータを記憶する。

【0021】図2は、URL「http://www.a.ne.jp/abc/」が与えられ、通信検索部121で取得されたHTMLテキストの一例を示す図である。

【0022】HTMLテキストの文法については「HTML入門 WWWページの作成と公開」(ローラ メリイ著 武舎広幸他訳 株式会社プレジデンスホール)などに記載されている。ここでは、本実施の形態に関係する部分を簡単に説明する。記述211~216は、ハイパーリンクを表す。記述216のハイパーリンクには画像データへのリンクが含まれている。

【0023】図3は、リンクされる画像データの一例を示す図である。

【0024】図4は、図3の画像データと図2のHTMLテキストからWebページが組み立てられて表示部109に表示されている状態をあらわす図である。

【0025】図5を参照して、記述211、216のハイパーリンクについて説明する。記述211は典型的なハイパーリンクの記述方法である。「」501は、開始タグで、ハイパーリンク記述の開始を表わし、「」504は終了タグで、ハイパーリンク記述の終了を表わす。開始タグ中の「intro. htm」502は、ハイパーリンクのリンク先であるURLである。「自己紹介」503はアンカー文字列で、表示部109には、図4に示す表示401のように下線などで強調表示される。記述216は、アンカー文字列の代わりに画像データが用いられている例である。「<A Href=http://w

ww. weather. com/>」505は開始タグ、「」509は終了タグ、「http://www. weather. com/」506は、リンク先のURLである。「」507は、表示する画像データのリンク先を示すイメージタグである。「天気」508は、画像データが表示できないテキスト専用ブラウザが使用されたとき画像データに替えて表示されるALT属性文字列である。

1. 1. 3 検索情報記憶部103

検索情報記憶部103は、磁気ディスクやメモリなどによって構成され、検索や検索結果の表示に必要な情報を記憶し、分野キーワード記憶部131と、トピックキーワード記憶部132と、検索起点記憶部133と、学習検索起点記憶部134と、検索履歴記憶部135とを含む。

1. 1. 3. 1 分野キーワード記憶部131

分野キーワード記憶部131は、ユーザによって入力部101から入力された検索キーワードを記憶する。この検索キーワードを分野キーワードと呼ぶことにする。図6は、分野キーワード記憶部131に記憶されている分野キーワードの一例を示す図である。「野球」601と「タイカース」602の2つの分野キーワードを記憶している。ここで2つの分野キーワードが記憶されているが、本発明は、これに限定されるものではなく、分野キーワード数は1でも、3でも実施可能である。

1. 1. 3. 2 トピックキーワード記憶部132

トピックキーワード記憶部132は、学習部108によって獲得された検索キーワードとその重要度とを記憶している。この検索キーワードをトピックキーワードと呼ぶことにする。

【0026】図7は、トピックキーワード記憶部132に記憶されているトピックキーワードとその重要度との一例を示す図である。トピックキーワード701~708と対応する重要度を示す数値711~718とが記載されている。ここでは、最大10個のトピックキーワードを記憶できることとする。ただし、本発明はこれに限定されるものではなくトピックキーワードの最大記憶数は20や30でも実施可能である。

1. 1. 3. 3 検索起点記憶部133

検索起点記憶部133は、ユーザによって入力部101から入力された検索起点のURLを記憶する。

【0027】図8は、検索起点記憶部133に記憶されている検索起点の一例を示す図である。検索起点のURL「http://www. a. ne. jp/abc/」801が記憶されている。ここで、検索起点は1つしか記憶していないが、本発明はこれに限定されるものではない。2つ以上記憶しても実施可能である。

1. 1. 3. 4 学習検索起点記憶部134

学習検索起点記憶部134は、学習部108によって獲

得された検索起点とその重要度とを記憶している。

【0028】図9は、学習検索起点記憶部134に記憶されている学習された検索起点とその重要度との一例を示す図である。検索起点901、902とそれぞれに対応する重要度911、912とが記載されている。ここでは、最大5個の検索起点を記憶できることとする。ただし、本発明はこれに限定されるものではなく検索起点の最大記憶数は10や20でも実施可能である。

1. 1. 3. 5 検索履歴記憶部135

10 検索履歴記憶部135は、学習部108によって学習された検索履歴を記憶する。1つの検索履歴は、URL、表示回数および非連続表示回数の3つ組で構成される。表示回数および非連続表示回数については、後で説明する。

【0029】図10は、検索履歴記憶部135に記憶されている検索履歴の一例を示す図である。各検索履歴のURL1001~1007とそれぞれに対応する表示回数1011~1017、非連続表示回数1021~1027との組が記載されている。ここでは、最大10個の検索履歴を記憶できることとする。ただし、本発明はこれに限定されるものではなく検索履歴の最大記憶数は20や30でも実施可能である。

1. 1. 4 関連語辞書104

関連語辞書104は、磁気ディスクなどによって構成され、関連語を記憶する。ここでの関連語とは、意味的に同じ単語、上位概念や下位概念の語、よく一緒に使用される語などである。関連語辞書104は、1つの単語に対して関連する関連語を参照できる構成となっており、各関連語には関連度が対応付けられて記載されている。

30 【0030】図11は、関連語辞書104の一例を示す図である。関連語辞書104は見出し語欄1101と、見出し語に対する関連語とその関連語の関連度とが記載された関連語欄1102とからなる。見出し語「野球」1110には、関連度「5」1112の関連語「ホームラン」1111と関連度「2」1114の関連語「ヒット」1113との2語が記載されている。ここで関連度は、見出し語と関連語との関連の強さを表わし、数値が大きいほど関連が強いことを表わす。

1. 1. 5 ハイパーリンク記憶部105

40 ハイパーリンク記憶部105は、RAMなどによって構成され、ハイパーリンクテーブルを記憶する。ハイパーリンクテーブルには、検索部106によって抽出されたハイパーリンクとその付随する情報や、表示情報作成部107の計算結果などが記載されている。図12は、ハイパーリンク記憶部105に記憶されているハイパーリンクテーブルの一例を示す図である。

【0031】ハイパーリンクテーブル1200は、URL欄1201と、取得フラグ欄1202と、リンク集フラグ欄1203と、抽出元番号欄1204と、アンカー文字列欄1205と、説明文字列欄1206と、ALT

属性文字列欄1207と、ページスコア欄1208と、リンクスコア欄1209と、表示スコア欄1210とを含んでいる。各欄1201~1209には、検索部106によって抽出されるハイパーリンクごとに対応したハイパーリンクに関する情報が記載され、表示スコア欄1210には、表示情報作成部107で計算された表示スコアが記載される。

【0032】URL欄1201には、検索開始時に第「0」行から検索起点記憶部133または学習検索起点記憶部134が記憶する検索起点のURLが記憶され、それ以降の行には、検索部106で抽出されたハイパーリンクのリンク先のURLが記載される。

【0033】取得フラグ欄1202には、取得フラグが記載される。取得フラグには、URL欄1201に記載されたURLに対応するHTMLテキストがネットワークから取得されていないときに「F」、取得されているときに「D」、これから取得することが決定されたときに「T」の値がそれぞれ記載される。

【0034】リンク集フラグ欄1203には、リンク集フラグが記載される。リンク集とは、ある基準に沿って多くのWebページへのハイパーリンクを記述したWebページのことである。URL欄1201に記載のURLで特定されるWebページがリンク集と予測できるときに「T」、予測できないときに「F」の値が記載される。

【0035】抽出元番号1204には、ハイパーリンクの抽出元HTMLテキストを示す抽出元番号が記載される。ハイパーリンクはハイパーリンクテーブル1200に登録（記載）されたURLに対応するHTMLテキストから抽出される。よって、その抽出元URLが登録されているハイパーリンクテーブル1200の行番号を抽出元番号として記載する。例えば、URL「http://www.a.ne.jp/abc/」に対応するHTMLテキストを抽出元URLとする抽出元番号は「0」となる。ただし、検索開始時には、ハイパーリンクテーブル1200には検索起点記憶部133に記憶されているURLが登録される。この際、抽出元番号に「-1」の値を記載する。

【0036】なお、行番号は、ハイパーリンクテーブル1200のURL欄1201の各URLの左側に示している。

【0037】アンカー文字列欄1205には、ハイパーリンクのアンカー文字列が記載される。

【0038】説明文字列欄1206には、ハイパーリンクの説明文字列が記載される。説明文字列は、ハイパーリンクの抽出元がリンク集のときだけ抽出されるハイパーリンクの後に記述される文字列である。説明文字列の終端は、次のハイパーリンクの先頭、HTMLテキストの終端、あるいは一定の文字列長を超えたときのいずれかとされる。説明文字列は、HTMLタグを含まな

い。

【0039】ALT属性文字列欄1207には、ALT属性文字列が記載される。ハイパーリンク中にイメージタグが記述され、その中にALT属性文字列が記述されていれば、この項目に記載される。

【0040】ページスコア欄1208には、ページスコアが記載される。ページスコアは、検索部106によってURLに対応するHTMLテキストがネットワークを介して取得されたときに計算される。HTMLテキスト中に出現した分野キーワード、トピックキーワード、関連語の出現回数とHTMLテキストのファイルサイズとから計算され、そのHTMLテキストの検索目的合致度を表わす。詳細は後述する。

【0041】リンクスコア欄1209には、リンクスコアが記載される。リンクスコアは、検索部106によってハイパーリンク中のアンカー文字列、説明文字列に出現した検索キーワード、トピックキーワード、関連語の出現回数から計算され、ハイパーリンクのリンク先のHTMLテキストの検索目的合致度の予測値を表わす。

【0042】表示スコア欄1210には、表示スコアが記載される。表示スコアは、表示情報作成部107によって計算される。

1. 1. 6 検索部106

検索部106は、入力部101からのユーザの指示により起動され、検索情報記憶部103に記憶されている情報に基づいて検索を行う。サブモジュールとして、検索制御部161と、検索初期化部162と、HTMLテキスト取得部163と、HTMLテキスト評価部164と、ハイパーリンク抽出部165と、ハイパーリンク評価部166と、検索終了判定部167と、ハイパーリンク選択部168とを有する。

1. 1. 6. 1 検索制御部161

検索制御部161は、検索処理全体の制御を行う。検索処理には、

- (1) 初期化处理
- (2) HTML取得処理
- (3) HTMLテキスト評価処理
- (4) ハイパーリンク抽出処理
- (5) ハイパーリンク評価処理
- (6) 終了判定
- (7) ハイパーリンク選択処理

とを含み、(2)から(7)が繰り返して実行される。終了判定で、終了が判定されると検索は終了する。

1. 1. 6. 2 検索初期化部162

検索初期化部162は、初期化处理を以下の手順で行う。

(1) ハイパーリンク記憶部105に記憶されているハイパーリンクテーブルをクリアする。

(2) 学習検索起点記憶部134に記憶されている検索起点のURLをハイパーリンクテーブル1200のUR

10

20

30

40

50

L欄1201の第「0」行から登録する。学習検索起点記憶部134に検索起点が記憶されていない場合、検索起点記憶部133に記憶されている検索起点のURLを登録する。

(3) 取得フラグ欄1202には「T」を、リンク集フラグ欄1203には「F」をそれぞれセットする。

(4) 抽出元番号欄1204には、学習検索起点記憶部134または検索起点記憶部133からURLが複写されたことを表す「1」をセットする。

【0043】検索起点記憶部133に図8に示す検索起点のURL801が記憶され、学習検索起点記憶部134に検索起点が記憶されていないとき、検索初期化部162は、図12に示したようにハイパーリンクテーブル1200のURL欄1201、取得フラグ欄1202、抽出元番号欄1203の各値を第「0」行に登録する。

1. 1. 6. 3 HTMLテキスト取得部163
HTMLテキスト取得部163は、以下のHTML取得処理を行う。

(1) ハイパーリンクテーブル1200から、取得フラグ欄1202に「T」が記載されたURL欄1201に記載されたURLを抽出する。

(2) 通信検索部121に抽出したURLを引き渡す。

(3) 通信検索部121が引き渡されたURLに対応するHTMLテキストをネットワークを介して取得し、通信バッファ122がそのHTMLテキストを記憶すると、取得フラグ欄1202に「D」をセットする。

【0044】図12に示すようなハイパーリンクテーブル1200がハイパーリンク記憶部105に記憶されているとき、URL「http://www.a.ne.jp/abc/」に対応するHTMLテキストが取得され、図13に示すようなハイパーリンクテーブル130*

*0に更新される。また、図17に示すハイパーリンクテーブル1700の状態のとき、URL「http://www.a.ne.jp/abc/baseball.htm」に対応するHTMLテキストが取得され、図18に示すようなハイパーリンクテーブル1800に更新される。さらに、図23に示すハイパーリンクテーブル2300の状態のとき、URL「http://www.a.ne.jp/abc/link.htm」に対応するHTMLテキストが取得され、図24に示すようなハイパーリンクテーブル2400に更新される。

1. 1. 6. 4 HTMLテキスト評価部164

HTMLテキスト評価部164は、以下の手順で取得したHTMLテキストのページスコアを計算し、ページスコア欄1208に登録する。

(1) HTMLテキスト中に分野キーワード記憶部131に記憶されている分野キーワードの出現する回数を計数する。

(2) HTMLテキスト中にトピックキーワード記憶部132に記憶されているトピックキーワードの出現する回数を計数する。

(3) HTMLテキスト中に分野キーワード記憶部131に記憶されている分野キーワードの関連語の出現する回数とその関連語との関連度の積和を算出する。たとえば、関連度「2」の関連語が3回、関連度「3」の関連語が1回出現すれば、 $3 \times 2 + 1 \times 3 = 9$ が算出される。

(4) HTMLテキストからHTMLタグを取り除いたファイルサイズを算出する。

(5) ページスコアは式(2)で計算する。

【0045】

【数2】

$$\text{ページスコア} = \frac{\begin{array}{l} \text{分野キーワード出現回数} \times 10 \\ + \text{トピックキーワード出現回数} \times 5 \\ + \text{関連語の出現回数と関連度の積和} \end{array}}{\text{HTMLテキストのHTMLタグを取り除いたファイルサイズ}} \dots (2)$$

ここで、図2に示すHTMLテキスト201が取得され、分野キーワード記憶部131に図6に示す分野キーワードが記憶され、トピックキーワード部132に図7に示すトピックキーワードとその重要度が記憶され、関連語辞書104に図11に示す見出し語と関連語とが記憶されている場合、ページスコアは以下のように計算される。

【0046】図2と図6とから、HTMLテキスト201には、分野キーワード「野球」1110が1回出現しており、分野キーワード出現回数「1」が計数される。

【0047】図2と図7とから、HTMLテキスト201には、トピックキーワードが出現しておらず、トピックキーワード出現回数「0」が計数される。

【0048】図2と図6と図11とから、HTMLテキ

スト201には、関連語が出現しておらず、関連語の出現回数と関連度の積和「0」が抽出される。

【0049】HTMLテキスト201からHTMLタグを取り除いたファイルサイズが200バイトであると、ページスコアは式(2)により、 $(1 \times 10 + 0 \times 5 + 0) / 200 = 0.05$ と計算される。図14は、ページスコア「0.05」が登録されたハイパーリンクテーブル1400の状態を表す。

【0050】また、図19に示すHTMLテキスト1901が取得された時には、以下のようにページスコアが計算される。分野キーワード「野球」601が1回出現することが計数され、トピックキーワード「半神」701、「シャイアンツ」702、「カーフ」703、「ヤグルト」704、「ヘイスターズ」705が各1回、合

計5回出現することが計数される。HTMLテキスト1901のファイルサイズが105バイトとすると、 $(1 * 10 + 5 * 5) / 105 = 0.33$ と計算され、図20に示すようにページスコア「0.33」が登録される。

【0051】また、図25に示すHTMLテキスト2501が取得された時には、HTMLテキスト2501に、分野キーワード「野球」601が2回、「半神」602が1回の合計3回出現していることが計数され、トピックキーワード「タイカース」701、「プロ」706が各1回の合計2回出現していることが計数され、関連語「ホームラン」1111が1回出現していることが計数される。関連語「ホームラン」1111の関連度は「5」なので、積和は $1 * 5 = 5$ が計算される。ファイルサイズが400バイトとすると $(3 * 10 + 2 * 5 + 5) / 400 = 0.11$ が計算され、図26に示すようにページスコア「0.11」が登録される。

1. 1. 6. 5 ハイパーリンク抽出部165

ハイパーリンク抽出部165は、以下のハイパーリンクの抽出処理を行う。

(1) 通信バッファ122に記憶された(取得された)HTMLテキストからハイパーリンクを抽出する。ハイパーリンクのURLとハイパーリンク記憶部105に記憶されているハイパーリンクテーブルのURL欄1201に記載のURLとを比較し、一致するものがなければ以下の処理を行う。

(2) ハイパーリンクテーブルのURL欄1201に抽出したハイパーリンクのURLを記載する。

(3) 取得フラグ1202に「F」を記載する。

(4) 抽出元番号欄1205に抽出先の行番号をセットする。

(6) アンカー文字列欄1205にアンカー文字列を記載する。

(7) 説明文字列欄1206に抽出元がリンク集なら説明文字列を記載する。

(8) ALT属性文字列欄1207にハイパーリンク中にイメージタグがあり、ALT属性文字列があれば、ALT属性文字列を記載する。

【0052】ハイパーリンク記憶部105に図14に示すようなハイパーリンクテーブル1400が記憶されている場合に取得したHTMLテキスト201が取得されたとき、ハイパーリンクの抽出処理が終わると、図15に示すようなハイパーリンクテーブル1500になる。ハイパーリンク211~216(図2)は第「1」行~第*

*「6」行に登録されている。

【0053】また、図20に示すようなハイパーリンクテーブル2000が記憶されている場合に、図19に示すHTMLテキスト1901が取得されているとき、ハイパーリンクの抽出処理が終わると、図21に示すようなハイパーリンクテーブル2100になる。ハイパーリンク1902は第「7」行に登録されるが、ハイパーリンク1903は、URLがすでにハイパーリンクテーブル2000の第「0」行に登録されているので登録されない。

【0054】また、図26に示すようなハイパーリンクテーブル2600が記憶されている場合に、図25に示すHTMLテキスト2501が取得されたとき、ハイパーリンクの抽出処理が終わると、図27に示すようなハイパーリンクテーブル2800になる。ハイパーリンク抽出元の第「7」行のリンク集フラグが「T」なので、登録された第「8」行~第「12」行のハイパーリンクには、説明文字列欄1205に説明文字列が抽出されている。

20 1. 1. 6. 6 ハイパーリンク評価部166

ハイパーリンク評価部166は、ハイパーリンクの評価処理を行う。評価処理は、リンク集フラグ欄1203のフラグの設定とリンクスコア欄1209のリンクスコアの計算とその計算値の設定とを行う。

【0055】リンク集フラグは、URL、アンカー文字列、ALT属性文字列のいずれかに「リンク」もしくは大文字小文字を同一視して「link」の文字が含まれていたら、リンク集フラグ欄1203に「T」をセットする。そうでなければ「F」をセットする。

30 【0056】リンクスコアは、以下の手順で計算する。

(1) アンカー文字列と説明文字列中とに分野キーワード記憶部131に記憶されている分野キーワードが出現する回数を計数する。

(2) アンカー文字列と説明文字列中とにトピックキーワード記憶部132に記憶されているトピックキーワードが出現する回数を計数する。

(3) アンカー文字列と説明文字列中とに分野キーワード記憶部131に記憶されている分野キーワードの関連語が出現する回数を計数する。

40 (4) 得られた計数結果から式(3)を用いてリンクスコアを計算し、その値をリンクスコア欄1209に登録する。

【0057】

【数3】

$$\begin{aligned} \text{リンクスコア} = & \text{分野キーワード出現回数} * 10 \\ & + \text{トピックキーワード出現回数} + \text{関連語出現回数} \dots (3) \end{aligned}$$

図15に示すハイパーリンクテーブル1500のとき、ハイパーリンクの評価処理を行うと、図16のようになる。ハイパーリンク1500の第「2」行のアンカー文

50 字列には分野キーワード「野球」601が含まれているのでリンクスコアは「10」となるが、他の第「1」行、第「3」~「6」行は「0」であるなお、このリン

クスコアは、リンク先のHTMLテキストの検索目的合致度の予測値を計算するものであるので、第「0」行については、計算されない。

【0058】また、図21に示すハイパーリンクテーブル2100のとき、ハイパーリンクの評価処理を行うと、図22のようになる。ここで、ハイパーリンクテーブル2200の第「7」行には、アンカー文字列に「リンク」を含むのでリンク集フラグ欄1203に「T」が設定される。

【0059】また、図27に示すハイパーリンクテーブル2700のとき、ハイパーリンクの評価処理を行うと、図28のようになる。

1. 1. 6. 7 検索終了判定部167

検索終了判定部167は、検索部106の検索の繰り返しの終了判定を行う。終了条件は、以下のようなものがある。

- (1) 一定個数以上のハイパーリンクを収集した。
- (2) 一定時間以上の検索を行った。
- (3) 一定値以上のリンクスコアを持つハイパーリンクが一定個数以上集まった。
- (4) ハイパーリンク選択部168で選択するハイパーリンクがなくなった。

【0060】以上のいずれかの又は組み合わせられた終了条件が満たされれば、検索終了判定部167は自動検索部106の処理を終了し、表示情報作成部107を起動する。ここでは、リンクスコアが1以上のハイパーリンクが3以上収集できたとき終了とする。図16、図22に示すハイパーリンクテーブル1600、2200では終了と判定しないが、図28に示すハイパーリンクテーブル2800では終了と判定する。

1. 1. 6. 8 ハイパーリンク選択部168

ハイパーリンク選択部168は、以下の基準で、ハイパーリンクテーブル中の取得フラグ欄1202のフラグが「F」のハイパーリンクから1つを選択する。

(基準1) リンク集フラグが「T」で、抽出元のページスコアがもっとも高いハイパーリンク。

(基準2) 基準1に該当するハイパーリンクがないとき、抽出元のハイパーリンクのリンク集フラグが「T」で、リンクスコアがもっとも高い外部サイトへのハイパーリンク。

【0061】なお、外部サイトか否かは注目しているハイパーリンクと抽出元番号で示される行とのURL欄1201の「//」と「/」とに挟まれている絶対パスが異なるとき外部サイトと判断する。

(基準3) 基準1、2に該当するハイパーリンクがないとき、リンクスコアがもっとも高いハイパーリンク。ただし、リンクスコア「0」のハイパーリンクは選択しない。

(基準4) 基準1、2、3に該当するハイパーリンクがないとき、抽出元のページスコアがもっとも高いハイ

パーリンク。ただし、ページスコア「0」から抽出されたハイパーリンクは選択しない。

【0062】ハイパーリンク選択部168は選択したハイパーリンクの、取得フラグ欄1202に「T」をセットする。

【0063】図16に示すハイパーリンクテーブル1600のとき、基準3によって第「2」行のハイパーリンクが選択され、図17に示すように取得フラグ欄1202に「T」がセットされる。

【0064】図22に示すハイパーリンクテーブル2200のとき、基準1によって第「7」行のハイパーリンクが選択され、図23に示すように取得フラグ欄1202に「T」がセットされる。

1. 1. 7 表示情報作成部107

表示情報作成部107は、サブモジュールとして、表示スコア検索部171と検索履歴検索部172と表示バッファ173とを有し、ハイパーリンクテーブルに登録されているハイパーリンクの表示スコアを計算し、検索履歴を検索して、表示回数が少なく表示スコアの高いものから順にハイパーリンクを並び替え、表示情報を作成する。

1. 1. 7. 1 表示スコア計算部171

表示スコア計算部171は、検索部106で作成され、ハイパーリンク記憶部105に記憶されているハイパーリンクテーブルの表示スコア欄1201の各表示スコアを以下の手順で計算する。

(1) ハイパーリンクの抽出元のリンク集フラグが「T」のとき、表示スコア欄1210に「500」を設定し、そうでなければ「0」を設定する。

(2) ハイパーリンクのリンクスコアが「0」のとき、表示スコアに「0」を設定し、リンクスコアが「1」以上なら、その値を表示スコアに加える。

【0065】図28に示すハイパーリンクテーブル2800のとき、表示スコア計算部171によって、図29に示すような、ハイパーリンクテーブル2900になる。

【0066】なお、本実施の形態では、表示スコアに設定する値を「500」または「0」としているけれども、この値は適宜修正可能である。

1. 1. 7. 2 検索履歴検索部172

検索履歴検索部172は、以下の手順でハイパーリンクテーブルから表示するハイパーリンクを抽出し、検索履歴記憶部135を参照して、ユーザの興味があるであろう順にハイパーリンクを並べ替え、表示バッファ173に表示情報を作成する。

(1) ハイパーリンクテーブルから表示スコアが1以上のものを抽出して、URL、アンカー文字列、説明文字列、表示スコアを表示バッファ173に表示情報として記載する。

(2) 抽出したハイパーリンクのURLで検索履歴記憶

部135を検索する。一致するURLが検索できたときは、対応する表示回数を表示バッファ173の表示情報表に記載する。検索できなかったときは、表示回数「0」を記載する。

(3) まず、表示回数で昇順に並び替える。

(4) 表示回数が同じハイパーリンクは、表示スコアで降順に並びかえる。

【0067】今、図29に示すようなハイパーリンクテーブル2900がハイパーリンク記憶部105に記憶され、検索履歴記憶部135に図10に示す検索履歴が記憶されているとき、検索履歴検索部172は、ハイパーリンクテーブル2900から、第「2」、「8」、「9」、「10」行のURL欄1201、アンカー文字列欄1205、説明文字列欄1206、表示スコア欄1210の各記載を抽出し、表示バッファ173の表示情報表に記載する。図30は、このときの表示バッファ173の内容を示している。表示情報表3000は、URL欄3001、アンカー文字列欄3002、説明文字列欄3003、表示スコア欄3004および表示回数欄3005とを有している。このとき、まだ表示回数欄3005への記載はない。

【0068】次に、検索履歴検索部172は、検索履歴記憶部135を検索し、表示回数を獲得する。図30に示す表示情報表3000の第「0」行のURLは、検索履歴記憶部135に一致するものがないので表示回数は「0」に設定される。他は、検索履歴記憶部135に記憶されている表示回数が記載される。

【0069】この結果、図30に示した表示情報表3000は、図31に示す表示回数欄3005に表示回数が記載された表示情報表3100に更新される。

【0070】最後に、表示回数と表示スコアで並び替えを行うと図32に示すような表示情報表3200になる。

1. 1. 7. 3 表示バッファ173

表示バッファ173は、RAMなどで構成され、図30～図32に示したように表示のために使用するハイパーリンクのアンカー文字列、説明文字列などの表示情報の一覧である表示情報表を記憶する。

1. 1. 8 学習部108

学習部108は、サブモジュールとして学習制御部181と、学習バッファ182と、キーワード学習部183と、検索起点学習部184と、検索履歴学習部185とを有し、表示情報作成部107が表示情報を作成すると起動される。

1. 1. 8. 1 学習制御部

学習部108が起動されると、学習制御部181は、学習バッファ182を初期化し、検索履歴の学習を行うため検索履歴学習部185を起動し、表示バッファ173の内容を表示部109に表示させる。図33は、図32に示した表示バッファ173の表示情報が表示部109

に表示された状態を表わす。表示部109は、一覧表示領域3301、ブラウザ領域3302、メッセージ表示領域3303、エージェント表示領域3304、○評価ボタン3305、×評価ボタン3306及び終了ボタン3307を有している。

【0071】一覧表示領域3301には、表示情報表のアンカー文字列欄3002に記載されたアンカー文字列が一覧表示される。表示されたアンカー文字列はユーザにより選択可能である。

【0072】学習制御部181は、アンカー文字列が入力部101のマウスを用いてクリックされると、アンカー文字列を反転表示し、対応する説明文字列をメッセージ表示領域3303に表示させる。併せて、エージェント表示領域3304に表示されるエージェント画像をアンカー文字列に対応する表示スコアが高いほど検索結果に自信のある表情に変化させる。図33では、「タイカース党」3313がクリックされた状態を表わしている。アンカー文字列がダブルクリックされると、対応するURLを通信部102に引き渡してWebページを取得し、ブラウザ領域3302に表示させる。また、ダブルクリックされたときは、その情報をキーワード学習部183に通知する。

【0073】○評価ボタン3305と×評価ボタン3306は、ユーザがブラウザ領域3302に表示されたWebページの善し悪し（ユーザの興味の有無）を以後のHTMLテキスト取得とエージェント画像とに反映するためのものである。○評価ボタン3305がクリックされると、エージェント表示領域3304に表示されたエージェント画像が喜んでいる画像を表示し、対応するメッセージをメッセージ表示領域3303に表示する。また、○評価ボタン3305がクリックされたことをキーワード学習部183に通知する。×評価ボタン3306がクリックされると、エージェント表示領域に表示されたエージェント画像が謝っている画像を表示し、対応するメッセージをメッセージ表示領域3303に表示する。また、×評価ボタン3306がクリックされたことをキーワード学習部183に通知する。

【0074】終了ボタン3307がクリックされると、検索結果の表示を終了し、キーワード学習部183を起動し、検索起点の学習を行うため検索起点学習部184を起動する。

1. 1. 8. 2 学習バッファ182

学習バッファ182は、RAMなどで構成され、表示部109に表示されたアンカー文字列がダブルクリックされた回数および、○評価ボタン3305、×評価ボタン3306がクリックされた回数を記録する。図34は学習バッファ182が学習制御部181によって初期化された状態を表わす。図34の第「0」～「3」行の各行は、図33のアンカー文字列3311～3314に対応する。各欄3401～3403には、ダブルクリックさ

れた回数、○評価ボタン3405がクリックされた回数、×評価ボタン3406がクリックされた回数がそれぞれ記録される。

1. 1. 8. 3 キーワード学習部183

キーワード学習部183は、学習制御部181から通知されたアンカー文字列のダブルクリック、○評価ボタン3405、×評価ボタン3406のクリックを学習バッファ182に記録する。

【0075】例えば図33を参照して、アンカー文字列「タイカース党」3313がダブルクリックされると、学習バッファ182の対応する第「2」行のダブルクリック欄3401の値を「1」増加させる。アンカー文字列「タイカース党」3313に対応するWebページが表示されているとき、○評価ボタン3305がクリックされると、学習バッファ182の対応する第「2」行の○評価ボタン欄3402の値を「1」増加させる。また、×評価ボタン3306がクリックされると、学習バッファ183の対応する第「2」行の×評価ボタン欄3403の値を「1」増加させる。

【0076】図35は、ユーザの操作によって変化した学習バッファ182の内容を表わす。

【0077】終了ボタン3307がクリックされると、以下の手順で、キーワード学習部183は、学習バッファ182と表示バッファ173とを参照して、トピックキーワード記憶部132のトピックキーワードの追加および更新をする。

(1) 表示バッファ173のアンカー文字列と説明文字列とからキーワード抽出を行う。キーワード抽出は形態素解析や字種(漢字、カタカナ、仮名)による切り出しなどによって行う。

(2) 抽出されたキーワードからトピックキーワード候補を抽出する。以下の条件のいずれかを満たすものを候補とする。

(条件1) ○評価ボタンのクリック回数が1以上の表示情報(アンカー文字列と説明文字列)から抽出されたキーワード

(条件2) ダブルクリック回数が1以上の複数の表示情報から抽出されたキーワード

(条件3) ダブルクリック回数が1以上の表示情報から抽出されたキーワードで、ダブルクリックされていない表示情報から抽出されていないキーワード

(3) ×評価ボタンのクリック回数が1以上の表示情報から抽出されたキーワードをトピックキーワード候補から削除する。

(4) 分野キーワードと一致するものをトピックキーワード候補から削除する。

(5) 残ったトピックキーワード候補にスコアを与える。スコアは、キーワードが抽出された表示情報の(ダブルクリック回数+○評価ボタンクリック回数*5)で与える。

(6) トピックキーワード候補のスコアの上位5つだけを残す。

(7) トピックキーワード記憶部132に記憶されているトピックキーワードとトピックキーワード候補を比較する。

(7-1) 一致したときは、トピックキーワード記憶部132に記憶されているトピックキーワードの重要度にトピックキーワード候補のスコアを加える。

(7-2) 不一致のときは、トピックキーワード候補をトピックキーワードとして登録する。トピックキーワードの重要度には、トピックキーワード候補のスコアを用いる。なお、トピックキーワード記憶部132に記憶する語数を超えるときは、重要度のもっとも低いトピックキーワードから削除する。

(8) トピックキーワード記憶部132に記憶されている、トピックキーワード候補とまったく一致しなかったトピックキーワードの重要度から一定値を減じる。

【0078】表示バッファ173が図32、学習バッファ182が図35、トピックキーワード記憶部132が図7のとき、以下のように学習が行われる。

【0079】まず、表示バッファ173のアンカー文字列と説明文字列からキーワード抽出を行う。ここでは形態素解析によって名詞をキーワードとして抽出する。なお、名詞以外の形容詞や動詞をキーワードとして抽出してもよい。図36(a)は、抽出されたキーワードを示す。

【0080】抽出されたキーワードからトピックキーワード候補を抽出する。以下の条件のいずれかを満たすものを候補とする。第「2」行のキーワードは条件1によりすべてトピックキーワード候補となる。また、キーワード「野球」は条件2によりトピックキーワード候補となる。また、「コーナー」「プロ」「リンク」「球団」「公式」が条件3によりトピックキーワード候補となる。図36(b)は、トピックキーワード候補を示す。

【0081】次に、×評価ボタンのクリック回数が1以上の表示情報から抽出されたキーワード(第「1」行)をトピックキーワード候補から削除する。図36(c)は、残ったトピックキーワード候補を示す。

【0082】分野キーワードと一致するものをトピックキーワード候補から削除する。ここでは「タイカース」が削除され図36(d)の状態になる。

【0083】残ったトピックキーワード候補にスコアを与える。スコアは、図36(e)のように、すべて6になる。トピックキーワード候補のスコアの上位5つだけを残す。ここでは全部同じなので最初の5つを選び、図36(f)になる。

【0084】トピックキーワード記憶部132に記憶されているトピックキーワードとトピックキーワード候補を比較する。

【0085】トピックキーワード候補「半神」はトピッ

クキーワード「半神」701に一致するので、トピックキーワードの重要度「23」711にトピックキーワード候補のスコア「6」を加え「29」となる。

【0086】不一致のときは、トピックキーワード候補をトピックキーワードとして登録する。トピックキーワードの重要度には、トピックキーワード候補のスコアを用いる。なお、トピックキーワード記憶部132に記憶する語数（ここでは「10」）を越えるときは、重要度のもっとも低いトピックキーワードを削除する。ここでは、「2」つしか新規登録するバッファがないので、重要度が低いトピックキーワード「ヤグルト」705と「プロ」706が削除される。図37は、この時点でのトピックキーワード記憶部132の内容を表わす。

【0087】最後に、トピックキーワード記憶部132に記憶されているトピックキーワード候補とまったく一致しなかったトピックキーワードの重要度から一定値「1」を減じる。図38、この時点でのトピックキーワード記憶部132の内容を表わす。

1. 1. 8. 4 検索起点学習部184

検索起点学習部184は、以下の手順で検索起点の学習を行う。

(1) 検索起点学習部184は、ハイパーリンクテーブルから以下の条件を満たすリンク集のURLを検索起点候補として抽出する。

(条件1) リンクスコアが「1」以上のハイパーリンクが抽出されたリンク集

(条件2) 学習バッファ182を参照し、ユーザにダブルクリックされ、×評価ボタンをクリックされていないアンカー文字列が抽出されたリンク集

なお、ユーザにダブルクリックされ、かつ○評価ボタンをクリックされたアンカー文字列に対応するページとしてもよい。

(2) 検索起点候補から検索起点記憶部133に記憶されている検索起点と一致するものを削除する。

(3) 検索起点候補のスコアとし、検索起点候補から抽出されたリンクスコアが「1」以上のハイパーリンク数を与える。

(4) 検索起点候補と学習検索起点記憶部134に記憶されている検索起点とを比較する。

(4-1) 一致したときは、学習検索起点記憶部134に記憶されている検索起点の重要度に検索起点候補のスコアを加える。

(4-2) 不一致のときは、検索起点候補を検索起点として学習検索起点記憶部134に登録する。検索起点の重要度には、検索起点候補のスコアを用いる。なお、学習検索起点記憶部134に記憶されているURLの数が所定の値を越えるときは、重要度のもっとも低い検索起点を削除する。

(5) 学習検索起点記憶部134に記憶されている、検索起点候補とまったく一致しなかった検索起点の重要度

から一定値を減じる。

【0088】ハイパーリンクテーブルが図28、学習バッファ182が図35、学習検索起点記憶部134が図9に示す内容のとき、以下のように学習する。

【0089】図28を参照して第「7」行のハイパーリンクが条件1を満たす。さらに、図35を参照して第「7」行は条件2も満たす。よって、ハイパーリンクテーブル2800の第「7」行のURL「http://www.a.ne.jp/abc/link.htm」が検索起点候補として抽出される。このURLは、検索起点記憶部133に記憶されている検索起点とは不一致なので、そのまま候補として残る。

【0090】検索起点候補のスコアとし、検索起点候補から抽出されたリンクスコアが1以上のハイパーリンク数「3」を与える。

【0091】次に検索起点候補と学習検索起点記憶部134に記憶されている検索起点とを比較すると、一致するので、学習検索起点記憶部134に記憶されている検索起点の重要度に検索起点候補のスコア「3」を加える。

【0092】最後に学習検索起点記憶部134に記憶されている、検索起点候補とまったく一致しなかった検索起点902の重要度912から一定値「1」を減じる。図39は、この時点での学習検索起点記憶部134の内容を表わす。重要度3901、3902の値が変更されている。

1. 1. 8. 5 検索履歴学習部185

検索履歴学習部185は、以下の手順で検索履歴の学習を行う。

(1) 検索履歴学習部185は、表示バッファ173に記憶されているハイパーリンクのURLと、検索履歴記憶部135に記憶されている検索履歴のURLとを比較する。

(1-1) 一致したときは、検索履歴の表示回数を

「1」増加し、非連続表示回数を「0」にセットする。

(1-2) 一致しないときは、表示バッファ173に記憶されているハイパーリンクのURLを検索履歴として登録する。表示回数に「1」、非連続表示回数に「0」をセットする。なお、検索履歴記憶部135に記憶される個数を越えているときは、非連続表示回数のもっとも多い検索履歴を削除する。

(2) 表示バッファ173に記憶されているハイパーリンクのURLと、まったく一致しなかった検索履歴の非連続表示回数を「1」増加させる。

【0093】表示バッファ173が図32、検索履歴記憶部135が図10に示す内容のとき、以下のように学習する。

【0094】表示バッファ173の第「1」行のURLは検索履歴1005のURLと、第2行のURLは検索履歴1003のURLと、第「3」行のURLは検索履歴

歴1004のいURLとそれぞれ一致する。そこで、検索履歴の表示回数を「1」増加し、非連続表示回数を「0」にセットする。

【0095】また、表示バッファ173の第「0」行のURLは検索履歴記憶部135のURLと一致しないので、表示回数を「1」、非連続表示回数を「0」とした、検索履歴を登録する。

【0096】最後に、表示バッファ173のURLと一致しなかった検索履歴1001、1002、1006、1007の非連続表示回数を「1」増加させる。図40は、この時点での検索履歴記憶部135の記憶内容を表わす。

1. 1. 9 表示部109

表示部109は、液晶ディスプレイなどで構成され、検索結果やWebページを表示する。

【0097】図41は、情報検索エージェント装置が起動された際の表示部109の表示内容を表わす。図33と同様の部分には同一符号を付して説明を省略する。検索ボタン4101は、ユーザがこれをクリックすると検索部106が起動される。

1. 2 動作

図42は、本実施の形態の動作を説明するフローチャートである。

【0098】情報検索エージェント装置は起動されると、表示部109に図41に示すような起動画面が表示される(S4201)。入力部101から検索の指示が与えられると(S4202)、検索部106が起動される。検索初期化部162は、ハイパーリンク記憶部105に記憶されているハイパーリンクテーブルをクリアし、学習検索起点記憶部134もしくは検索起点記憶部135が記憶されている検索起点を登録する初期化処理を行う(ステップS4203)。HTMLテキスト取得部163は、ハイパーリンクテーブルを参照して必要なHTMLテキストをネットワークを介して取得する(S4204)。HTMLテキスト評価部164は、取得したHTMLテキストの検索要求への合致度をキーワードを用いて評価する(S4205)。ハイパーリンク抽出部165は、取得したHTMLテキストからハイパーリンクを抽出する(S4206)。ハイパーリンク評価部166は、抽出されたハイパーリンクの検索要求への合致度及びリンク集の判定などの評価を行う(S4207)。検索終了判定部167は、繰り返し検索の終了を判定する(S4208)。検索を継続する場合は、次に取得するHTMLテキストを決定するため、ハイパーリンクテーブルに登録されたハイパーリンクの選択を行い(S4209)、制御をステップS4204へ移す。

【0099】検索を終了すると、表示情報作成部107が起動され、表示情報を作成する(S4210)。検索履歴学習部184は、作成された表示情報から、検索履歴を学習し検索履歴記憶部185を更新する(S421

1)。表示部109に、表示情報を参照して、検索結果としてアンカー文字列を一覧表示する(S4212)。

【0100】学習制御部181は、学習バッファ183をクリアして初期化する(S4213)。

【0101】アンカー文字列がダブルクリックされると(S4214)、キーワード学習部183は学習バッファ182のダブルクリック回数を「1」増加させて更新する(S4215)。○評価ボタン3305がクリックされると(S4216)、学習バッファ182の○評価ボタンクリック回数を「1」増加させ更新する(S4217)。×評価ボタン3306がクリックされると(S4218)、学習バッファ182の×評価ボタンクリック回数を「1」増加させ更新する(S4219)。

【0102】終了ボタン3407がクリックされると(S4220)、検索起点学習部184は、ハイパーリンクテーブルおよび学習バッファ182を参照して検索起点候補を抽出し、学習検索起点記憶部134の更新および登録を行う(S4221)。キーワード学習部183は、学習バッファ182と表示バッファ173を参照してトピックキーワード候補を抽出し、トピックキーワード記憶部132の更新および登録を行い(S4222)、S4201に戻る。

2 実施の形態2

2. 1 構成

図43は、本発明に係る情報検索エージェント装置の実施の形態2の構成図である。

【0103】この情報検索エージェント装置は、入力部101と、通信部102と、検索情報記憶部103と、関連語辞書104と、ハイパーリンク記憶部105と、自動検索部106と、表示情報作成部107と、学習部108と、表示部109と、検索情報管理部4301と、検索情報制御部4302とを備えている。なお、上記実施の形態1と同一の構成部分には同一の符号を付してその説明を省略し、本実施の形態固有の構成部分について説明する。

2. 1. 1 検索情報管理部4301

検索情報管理部4301は、磁気ディスクなどで構成され、検索情報記憶部103に複写するための検索情報を名前をつけて複数記憶している。図44は、検索情報管理部4301に記憶されている複数の検索情報の一例を示す図である。

【0104】ここでは、名前として「野球」4411、「競馬」4421、「料理」4431が付けられた3つの検索情報4410、4420、4430が記憶されている。各検索情報4410、4420、4430には、名前4411、4421、4431、分野キーワード4412、4422、4432、トピックキーワード4413、4423、4433、検索起点4414、4424、4434、学習された検索起点4415、4425、4435及び検索履歴4416、4426、443

6がそれぞれ記載されている。なお、検索履歴4416は、例えば図40に示したような内容である。

2. 1. 2 検索情報制御部4302

検索情報制御部4302は、情報検索エージェント装置が起動されたとき、検索情報管理部4301を参照して、検索情報に付けられた名前を表示部109に表示させる。図45は、検索情報管理部4301に図44に示す3つの検索情報4410、4420、4430が記憶されているときの起動時の表示部109の表示画面を示す図である。一覧表示領域3301には、検索情報の名前「野球」4501、「競馬」4502、「料理」4503が表示されている。ここで、表示された名前をユーザがクリックすることにより選択され、選択された検索情報の名前は反転表示される。図45では、「野球」4501が選択され、反転表示されている。ユーザによる検索ボタン4101がクリックされると、検索情報制御部4302は、選択された名前の検索情報を、検索情報管理部4301から読出し、検索情報記憶部103に複写する。例えば、検索情報4410が検索情報記憶部103の各記憶部131~135に記載される。

【0105】また、検索情報制御部4302は、検索結果が表示されている状態で終了ボタン3307（図33）がクリックされると、検索情報記憶部103に記憶されている検索情報を、起動時に選択された名前の検索情報として検索情報管理部4301に記載する。

（実施の形態3）

3. 1 構成

図46は、本発明に係る情報検索エージェント装置の実施の形態3の構成図である。

【0106】この情報検索エージェント装置は、入力部101と、通信部102と、検索情報記憶部103と、関連語辞書104と、ハイパーリンク記憶部105と、検索部106と、表示情報作成部107と、学習部108と、表示部109と、一括取得部4601と、Webページ記憶部4602とを備えている。

【0107】なお、上記実施の形態1と同一の構成部分には、同一の符号を付してその説明を省略し、本実施の形態固有の構成部分について説明する。

3. 1. 1 一括取得部4601

一括取得部4601は、表示情報作成部107で作成された表示情報のURLを順次、通信部102に引き渡す。通信部102はURLに対応するWebページを構成するHTMLテキストおよびイメージデータをネットワークを介して取得する。一括取得部4601は、通信部102で取得されたWebページ（実際にはHTMLテキストおよびイメージデータ）をWebページ記憶部4602に記憶させる。

【0108】一括取得を行うことにより、通信回線を切断し、取得保存したWebページを閲覧するようすることで、通信コストを削減することができる。

3. 2. 1 Webページ記憶部4602

Webページ記憶部4602は、磁気ディスクなどで構成され、一括取得部4601で通信部102を介して取得されたWebページを記憶している。

【0109】なお、上記実施の形態1~3では、ページスコア、リンクスコア、表示スコア、トピックキーワード候補のスコア、検索起点候補のスコアの各計算方法には所定の定数を用いた計算式を示したけれども、計算式はこれらに限るものではなく、また定数も適宜修正可能である。

【0110】また、上記実施の形態1~3では、通信検索部121で取得される情報をHTMLテキストとしたけれども、他のハイパーリンクが記述された例えばXML (Extensible Markup Language) テキストであってもよい。

4 実施の形態の変形例

4. 1 その1

上記実施の形態1~3においてハイパーリンク評価部166は、ハイパーリンクの評価を決定する際、分野キーワード、トピックキーワード、関連語によって評価を行っているが、この方法に限定されるものではない。ここで、HTMLテキストを提供するサイト情報を記憶するサイト情報記憶部を設けても良い。

【0111】図47は、サイト情報記憶部に記憶されている内容の一例を示す図である。ここで、サイト情報記憶部は、サイトのURL4701と通信速度4702とを記憶している。通信速度は単位時間あたりの平均送信データ量（バイト数）のことで、通信部102で取得されたデータのサイズを通信時間で割ることによって得られるもので、過去の通信履歴から作成される。図47を参照すると、例えば、URL「http://www.a.ne.jp/」のサイトの通信速度は、URL「http://www.b.co.jp/」のサイトの通信速度の2倍以上であり、同じ時間で倍以上の情報を得られることがわかる。このサイト情報を利用して通信速度の速いサイトのURLを持つハイパーリンクの評価値を上げ、より速く多くの情報を収集することが可能となる。通信速度の評価値への組み込みの具体的な方法の1つは、ハイパーリンクのURLのサイト部分とサイト情報記憶部のURLを比較し、一致したときに通信速度を一定値で割った値をハイパーリンクの評価値に加えることである。また、キーワードによって計算された評価値に通信速度を掛ける方法などがある。

4. 2 その2

【0112】また、サイト情報として、ユーザにとって有益な情報を提供しているかを記憶することもできる。図48は、サイト情報記憶部に記憶されている内容の一例を示す図である。ここで、サイト情報記憶部は、サイトのURL4801と表示回数4802とを記憶している。表示回数はこれまでにユーザによって閲覧された回

数を記録したものである。表示部109に表示された回数を計数することで作成される。また、単なる回数でなく表示時間で計数しても良い。また、起動時からの全表示回数もしくは表示時間でなく、最近の一定期間だけの表示回数もしくは表示時間を計数することによって、そのときのユーザの好みを反映させることができる。この表示回数あるいは表示時間の長いサイトは、ユーザにとって有益な情報を多く提供していると考えられる。このサイト情報を利用して有益な情報の多いサイトのURLを持つハイパーリンクの評価値を上げ、より速く有益な情報を収集することが可能となる。評価値への組み込みの具体的な方法の1つは、ハイパーリンクのURLのサイト部分とサイト情報記憶部のURLのサイト部分とを比較し、一致したときに表示回数を一定値で割った値を評価値に加えることである。また、キーワードによって計算された評価値に表示回数を掛ける方法などがある。

4. 3 その3

【0113】また、一定期間の過去の検索参照履歴を記憶する検索参照履歴記憶部を設け、これをハイパーリンクの評価値に反映させても良い。検索参照履歴とは、検索が行われたとき、表示部109に表示されたハイパーリンクの一覧で、参照されなかった回数を記録したものである。この回数も最近の一定期間だけを計数することによって、そのときのユーザの好みを反映させることができる。図49は、検索参照履歴記憶部に記憶されている検索参照履歴の一例を示す図である。URL4901に対して、表示部105に表示されたにも拘わらず参照されなかった回数4902が記憶される。この回数が多いほどユーザにとって不要な情報といえる。評価値への組み込みの具体的な方法の1つは、ハイパーリンクのURLと検索参照履歴記憶部のURLを比較し、一致したときに参照されなかった回数を一定値で割った値を評価値から減じることである。また、この検索参照履歴記憶部では参照されなかった回数のみを記憶しているが、参照した回数も記録して、その回数を一定値で割った値を評価値に加えることも可能である。

4. 4 その4

【0114】また、ハイパーリンクの評価値を決定する際、探索範囲を広げるために、一定の割合でランダムにハイパーリンクを選択することも可能である。

4. 5 その5

また、上記実施の形態1～3では、検索において、1つのハイパーリンクを選択し、対応するHTMLテキストを取得する繰り返しとしているが、これを複数のハイパーリンクを選択し、それらすべての対応するHTMLテキストを取得することを1回の検索としても実施可能である。また、このとき、複数のハイパーリンクの決定方法を用意して、それぞれの基準でハイパーリンクを選択することも可能である。

4. 6 その6

【0115】ハイパーリンクの選択基準にリンク集か否かを主に用いたけれども、HTMLテキストの種類「目次」、「コンテンツ」、「リンク集」なども判定し、表示やハイパーリンクの選択に利用してもよい。

4. 7 その7

また、上記実施の形態1～3では、表示の際、表示バッファ173に記憶されているすべての情報を表示したけれども、これを検索履歴記憶部135に記憶しているURLと一致するハイパーリンクだけに限定してもよい。このように限定することにより、過去に良く参照したページの一覧表示を実現できる。また、一定の適合度（リンクスコア）以上のハイパーリンクのみ表示しても良い。また、表示の際、ハイパーリンクに対する適合度を同時に表示してもよい。また、HTMLテキストの内容である種別を表示してもよい。図50は、一覧表示領域3301に適合度と種別も同時に表示した例である。ここで適合度は10以上のとき「◎」、5以上のとき「○」で表示している。

【0116】4. 8 その8

また、適合度や種別で並べ替えるのではなく、各HTMLテキストのリンク関係を反映した表示としてもよい。図51は、一覧表示領域3301にHTMLテキストのリンク関係と適合度とを表示している例である。ハイパーリンクがツリー状に表示されている。上記変形例その7と同様、各ハイパーリンクの適合度は10以上のとき「◎」、5以上のとき「○」で表示している。

4. 9 その9

なお、上記実施の形態1～3において、ユーザが検索情報を登録、削除、変更するための検索情報変更部を設けるようにしてもよい。ユーザは、検索情報として、検索情報の名前、分野キーワード、検索起点のURLを登録することで、必要な検索情報を必要なだけ用意することができる。これにより、Webページを閲覧中にユーザが有効なキーワードを見つけた場合、そのキーワードを分野キーワードとして検索情報部変更部から追加することができる。

【0117】また、使用しなくなった検索情報を削除することで、検索情報の記憶領域を削減することができる。

4. 10 その10

上記実施の形態1～3では、トピックキーワードをアンカー文字列や説明文字列から抽出したけれども、従来のようにHTMLテキスト全体からトピックキーワードを抽出するようにしてもよい。

4. 11 その11

また、上記実施の形態1～3では、キーワード学習部183は、一覧表示されているハイパーリンクが選択されたときに起動されているが、選択された後一定時間閲覧された際に起動されても良い。ユーザがハイパーリンク

を選択し、対応するWebページを表示した結果、有益な情報であれば、長く閲覧し、不要な情報であれば閲覧を即やめることが考えられる。そこで、ハイパーリンクの選択後、一定時間対応するWebページを表示しているときに限り、キーワード学習部183が学習バッファ182のダブルクリック回数の更新を行うようにして、よりの確にユーザの好みを反映することができる。

【0118】4.12 その12

上記実施の形態1～3の情報検索エージェント装置は、サーバー・クライアントの形態で実現することも可能である。クライアント側は入力部101、表示部109、検索情報記憶部103、検索情報管理部4301、検索情報制御部4302を備え、それ以外の構成要素はサーバー側が備える。サーバーは、必要な検索情報をクライアント側から取得して検索処理を行う。また、サーバーは検索結果をクライアントの表示部109に表示する。検索情報は個人の嗜好を表す個人情報であり、各個人が所有するパソコンなどのクライアントマシンで管理するのが望ましい。また、検索処理は、処理能力が高い共用のサーバーマシンを使用することによって高速化することができ

【0119】4.13 その13

上記実施の形態1～3の情報検索エージェント装置は、リンク集フラグを利用してリンク集を優先的に取得していた。しかし、XMLなどのハイパーリンク記述形式において、リンク先のページが多くのハイパーリンクを含むことを表す属性を記述できるとき、この属性をリンク集フラグの決定に利用しても良い。例えばハイパーリンクの開始タグを「」と記述し、このうちの「num=10」がハイパーリンク先のページに含まれるハイパーリンク個数を表すとする。この個数が所定の値以上であればリンク集フラグを「T」としてもよい。またリンク集フラグをリンク集合スコアとし、このハイパーリンク個数を記憶して、ハイパーリンク個数の多いページから優先的に取得するようにしても良い。

4.14 その14

【0120】なお、本発明に係る情報検索エージェント装置は、上記実施の形態1～3において、それぞれ図1、図43及び図46に構成図として示したけれども、各構成部分の機能を発揮するプログラムを、フロッピーディスク等のコンピュータ読み取り可能な記録媒体に記録して移送する、あるいは、通信回線を介して移送することにより、独立した他のコンピュータシステムで容易に実施することができる。

【0121】

【発明の効果】以上説明したように、本発明によれば、ネットワーク及びシステム上のアドレスで指定されるデータを検索する情報検索エージェント装置であって、アドレスと検索開始の指示とを受け取るか又はアドレスの通

知を受けると、当該アドレスで指定されるデータを取得するデータ取得手段と、取得された複数のデータ要素からなるデータを記憶するデータ記憶手段と、前記データ記憶手段に記憶されているデータを解析し、データ要素に含まれる他のデータへのリンク先アドレスと当該アドレスに関連する文字列とを抽出する抽出手段と、抽出された文字列に基づいてリンク先アドレスで指定されるデータの選択適性の評価値を計算する計算手段と、評価テーブルに、文字列と、計算された評価値とをリンク先アドレスごとに記載する評価テーブル記載手段と、前記評価テーブルに記載された評価値が最高のリンク先アドレスを前記データ取得手段に通知するデータ取得制御手段と、前記データ取得制御手段の処理を前記評価テーブルの状況に応じて禁止する禁止手段と、前記評価テーブルに記載された文字列を選択する選択手段と、前記選択手段で選択された文字列を表示画面に一覧表示させる一覧表示制御手段とを備えることとしている。このような構成によって、文字列に関連するリンク先アドレスで指定されるデータを予め取得することなく、表示画面に文字列を一覧表示させるので、無駄なデータ取得のための通信コストを削減することができ、また、評価テーブルの状況に応じて検索を終了するので、データの記憶領域も少なくてもよい。

【0122】また、前記表示画面に表示された文字列の選択をユーザから受け取ると、対応するリンク先アドレスを前記データ取得手段に通知するデータ取得通知手段と、前記データ取得通知手段によるアドレスの通知の後、前記データ記憶手段にデータが記憶されると、前記抽出手段の処理を禁止して、記憶されているデータを解析し、表示画面に表示させるデータ表示制御手段とを備えることとしている。このような構成によって、ユーザが必要とするデータを容易に表示させることができる。

【0123】また、ユーザからキーワードの入力を受け付けて記憶しているキーワード記憶手段と、見出し語と、その見出し語に関連する関連語と、その関連語の見出し語に対する関連度とが登録された関連語辞書とを備え、前記文字列は、リンク先アドレスで指定されるデータの内容を示す文字列であり、文字列には説明文が付されている場合があり、前記計算手段は、前記文字列及び説明文に含まれる単語が前記キーワード記憶手段に記憶されているキーワードと一致するとき、評価値を大きくするキーワード評価部と、前記文字列及び説明文に含まれる単語が前記キーワード記憶手段に記憶されているキーワードを前記関連語辞書の見出し語とする関連語と一致するとき、その関連度に応じて評価値を大きくする関連語評価部とを有することとしている。このような構成によって、データの選択適性がユーザの興味あるキーワードによって評価されるので、的確なデータの選択予測が可能となり、無駄なデータ検索をすることが防止され、通信コストの削減が図れる。

【0124】また、前記表示画面に表示された文字列の選択をユーザから受け付けると、当該文字列を形態素解析し、所定の条件により、当該文字列に含まれる単語をトピックキーワードに決定するトピックキーワード決定手段と、前記決定されたトピックキーワードを記憶するトピックキーワード記憶手段とを備え、前記計算手段は、前記文字列及び説明文に含まれる単語が前記トピックキーワード記憶手段に記憶されているトピックキーワードと一致するとき、評価値を大きくするトピックキーワード評価部を有することとしている。このような構成によって、ユーザの表示データの選択を受けて、ユーザの興味を反映させつつ、データの選択適性を評価しているので、よりの確なデータ検索をすることができる。

【0125】また、前記データ取得制御手段は、前記評価テーブルに記載されたリンク先アドレスと文字列とのいずれかに「リンク」又は「link」の文字列が含まれているか否かを判断する判断部と、前記判断部で含まれていると判断されたとき、当該リンク先アドレスを評価値が最高のリンク先アドレスに優先して前記データ取得手段に通知する優先通知部とを有することとしている。このような構成によって、リンク先アドレス等に「link」等の単語を含むデータはリンク集であることが予測され、そのデータからリンクされる他のデータを多数有しているので、効果的なデータ検索ができる。

【0126】また、前記一覧表示制御手段によって文字列が表示されたとき、文字列ごとに表示回数を計数する計数手段と、前記計数手段で計数された表示回数を対応するリンク先アドレスごとに記憶する検索履歴記憶手段とを備え、前記選択手段は、前記検索履歴記憶手段に記憶されている表示回数の多いリンク先アドレスに対応する文字列を劣後する劣後部を有することとしている。このような構成によって、何度も同じデータを表示させることを防止することができ、新しいデータを優先的に表示できる。

【0127】また、前記データ表示制御手段でデータが表示画面に表示されているときに、ユーザから興味の有無に関するボタン操作を受け付ける受付手段と、前記受付手段で興味「有」を受け付けたとき選択されている文字列に対応するリンク先アドレスを検索起点として登録する検索起点登録手段と、前記受付手段で興味「無」を受け付けたとき選択されている文字列に対応するリンク先アドレスが検索起点登録手段で登録されているとき、その検索起点であるリンク先アドレスを削除する検索起点削除手段とを備え、前記データ取得手段は、検索開始の指示を受けたとき、検索起点登録手段に登録されている当該アドレスで指定されるデータを取得することとしている。このような構成によって、ユーザが検索起点のアドレスを入力しなくても、過去の履歴から検索開始のデータを決定でき、またユーザの興味の有無を反映させた新しいデータの検索ができる。

【0128】また、前記トピックキーワード記憶手段に記憶されているトピックキーワードには、重要度が付されており、前記データ表示制御手段でデータが表示画面に表示されているときに、ユーザから興味の有無に関するボタン操作を受け付ける受付手段を備え、前記トピックキーワード決定手段は、前記受付手段で興味「有」を受け付けたとき、対応する文字列から決定されたトピックキーワードの重要度を高くし、興味「無」を受け付けたとき、対応する文字列から決定されたトピックキーワードの重要度を低くする重要度評価部を有し、前記トピックキーワード評価部は、この重要度を考慮することとしている。このような構成によって、ユーザの興味の有無を反映させたデータ検索ができる。

【0129】また、データ内容の分野別に名前を付し、検索開始時のアドレスと、キーワードと、トピックキーワードと、前記一覧表示制御手段での表示履歴とを含む検索情報を記憶する検索情報記憶手段と、検索終了時に前記検索情報を更新する検索情報更新手段とを備え、前記一覧表示制御手段は、ユーザから検索開始の指示を受けると、前記名前を表示画面に一覧表示させる初期表示部を有し、前記データ取得制御手段は、前記名前の選択を受けると、対応する検索開始時のアドレスを前記データ取得手段に通知する初期通知部を有することとしている。このような構成によって、ユーザが選択した分野別のデータの検索が容易にできる。

【0130】また、前記計算手段は、前記データ取得手段がデータを取得するのに要した時間を計時し、取得したデータ量とから送信速度を計算する送信速度計算部と、前記送信速度計算部で計算された送信速度を同一サイトのアドレスごとに記憶する送信速度記憶部と、送信速度の速いサイトの評価値を大きくする送信速度評価部とを有することとしている。このような構成によって、短時間で取得できるサイトを優先するので、通信コストを削減できる。

【0131】また、前記計算手段は、前記データ表示制御手段でデータが表示画面に表示されたとき、その表示回数を計数するデータ表示回数計数部と、計数された表示回数を同一サイトのアドレス毎に記憶するデータ表示回数記憶部と、前記データ表示回数の多いサイトの評価値を大きくする表示回数評価部とを有することとしている。このような構成によって、ユーザの興味のあるサイトからのデータの検索が容易にできる。

【0132】更に、本発明はネットワークやシステム上の所定のアドレスで指定されるデータを検索する情報検索エージェント装置であって、アドレスと検索開始の指示とを又はアドレスの通知を受けると、当該アドレスで指定されるデータを取得するデータ取得手段と、取得された複数のデータ要素からなるデータを記憶するデータ記憶手段と、前記データ記憶手段に記憶されているデータを解析し、データ要素に含まれる他のデータへのリン

ク先アドレスと当該アドレスで指定されるデータに関連する文字列とを抽出する抽出手段と、抽出された文字列に基づいてリンク先アドレスで指定されるデータの選択適性の評価値を計算する計算手段と、評価テーブルにリンク先アドレスと、文字列と、計算された評価値とをリンク先アドレスごとに記載する評価テーブル記載手段と、前記評価テーブルに記載された評価値が最高のリンク先アドレスを前記データ取得手段に通知するデータ取得制御手段と、前記データ取得制御手段の処理を前記評価テーブルの状況に応じて禁止する禁止手段と、前記評価テーブルに記載された所定の文字列を選択する選択手段と、前記選択手段で選択された文字列に対応するリンク先アドレスで指定されるデータを順次取得する一括データ取得手段と、前記一括データ取得手段で取得されたデータを記憶する一括データ記憶手段とを備えることとしている。このような構成によって、ユーザが興味を持つデータを予測して取得しておくので、希望のデータを時間を要せずして表示させることができる。

【0133】更に、本発明はコンピュータ読み取り可能な記録媒体であって、ネットワークやシステム上の所定のアドレスで指定されるデータを検索する情報検索エージェント装置の、アドレスと検索開始の指示とを又はアドレスの通知を受けると、当該アドレスで指定されるデータを取得するデータ取得手段と、取得された複数のデータ要素からなるデータを記憶領域に書き込むデータ書込手段と、前記データ記憶領域に記憶されているデータを解析し、データ要素に含まれる他のデータへのリンク先アドレスと当該アドレスで指定されるデータに関連する文字列とを抽出する抽出手段と、抽出された文字列に基づいてリンク先アドレスで指定されるデータの選択適性の評価値を計算する計算手段と、評価テーブルにリンク先アドレスと、文字列と、計算された評価値とをリンク先アドレスごとに記載する評価テーブル記載手段と、前記評価テーブルに記載された評価値が最高のリンク先アドレスを前記データ取得手段に通知するデータ取得制御手段と、前記データ取得制御手段の処理を前記評価テーブルの状況に応じて禁止する禁止手段と、前記評価テーブルに記載された所定の文字列を選択する選択手段と、前記選択手段で選択された文字列を表示画面に一覧表示させる一覧表示制御手段との各手段の機能を発揮させるプログラムを記録することとしている。このような構成によって、情報検索の効率が向上した情報検索エージェント装置としてコンピュータを利用することができる。

【図面の簡単な説明】

【図1】本発明に係る情報検索エージェント装置の実施の形態1の構成図である。

【図2】上記実施の形態1の通信検索部で取得されたHTMLテキストの一例を示す図である。

【図3】上記実施の形態のHTMLテキストにリンクさ

れる画像データの一例を示す図である。

【図4】上記実施の形態の表示部に図2に示したHTMLテキストを表示させた表示例を示す図である。

【図5】上記実施の形態で用いられるHTMLテキストのハイパーリンクの説明図である。

【図6】上記実施の形態の分野キーワード記憶部の記憶内容の一例を示す図である。

【図7】上記実施の形態のトピックキーワード記憶部の記憶内容の一例を示す図である。

【図8】上記実施の形態の検索起点記憶部の記憶内容の一例を示す図である。

【図9】上記実施の形態の学習検索起点記憶部の記憶内容の一例を示す図である。

【図10】上記実施の形態の検索履歴記憶部の記憶内容の一例を示す図である。

【図11】上記実施の形態の関連語辞書の記憶内容の一例を示す図である。

【図12】上記実施の形態のハイパーリンク記憶部の起動直後の記憶内容の一例を示す図である。

【図13】上記実施の形態のハイパーリンク記憶部の記憶内容の一例を示す図である。

【図14】上記実施の形態のハイパーリンク記憶部の記憶内容の一例を示す図である。

【図15】上記実施の形態のハイパーリンク記憶部の記憶内容の一例を示す図である。

【図16】上記実施の形態のハイパーリンク記憶部の記憶内容の一例を示す図である。

【図17】上記実施の形態のハイパーリンク記憶部の記憶内容の一例を示す図である。

【図18】上記実施の形態のハイパーリンク記憶部の記憶内容の一例を示す図である。

【図19】上記実施の形態の通信検索部121で新たに取得されたHTMLテキストを示す図である。

【図20】上記実施の形態のハイパーリンク記憶部の記憶内容の一例を示す図である。

【図21】上記実施の形態のハイパーリンク記憶部の記憶内容の一例を示す図である。

【図22】上記実施の形態のハイパーリンク記憶部の記憶内容の一例を示す図である。

【図23】上記実施の形態のハイパーリンク記憶部の記憶内容の一例を示す図である。

【図24】上記実施の形態のハイパーリンク記憶部の記憶内容の一例を示す図である。

【図25】上記実施の形態の通信検索部121で新たに取得されたHTMLテキストを示す図である。

【図26】上記実施の形態のハイパーリンク記憶部の記憶内容の一例を示す図である。

【図27】上記実施の形態のハイパーリンク記憶部の記憶内容の一例を示す図である。

【図28】上記実施の形態のハイパーリンク記憶部の記

憶内容の一例を示す図である。

【図29】上記実施の形態のハイパーリンク記憶部の記憶内容の一例を示す図である。

【図30】上記実施の形態の表示バッファの記憶内容の一例を示す図である。

【図31】上記実施の形態の表示バッファの記憶内容の一例を示す図である。

【図32】上記実施の形態の表示バッファの記憶内容の一例を示す図である。

【図33】上記実施の形態の表示部の表示内容の一例を示す図である。 10

【図34】上記実施の形態の学習バッファの初期化された状態を示す図である。

【図35】上記実施の形態の学習バッファの変化した内容を示す図である。

【図36】(a)は、上記実施の形態のキーワード学習部によって抽出されたキーワードを示す図である。

(b)～(f)は、上記実施の形態のキーワード学習部によって抽出されたトピックキーワード候補を示す図である。 20

【図37】上記実施の形態のトピックキーワード記憶部の記憶内容の一例を示す図である。

【図38】上記実施の形態のトピックキーワード記憶部の記憶内容の一例を示す図である。

【図39】上記実施の形態の学習検索起点記憶部の記憶内容の一例を示す図である。

【図40】上記実施の形態の検索履歴記憶部の記憶内容の一例を示す図である。

【図41】上記実施の形態の情報検索エージェント装置が起動された際の表示部の表示内容を示す図である。 30

【図42】上記実施の形態の動作を説明するフローチャートである。

【図43】本発明に係る情報検索エージェント装置の実施の形態2の構成図である。

【図44】上記実施の形態の検索情報管理部に記憶されている記憶内容の一例を示す図である。

【図45】上記実施の形態の起動時の表示部の表示内容の一例を示す図である。

【図46】本発明に係る情報検索エージェント装置の実施の形態3の構成図である。

【図47】本発明に係る情報検索エージェント装置の上記実施の形態の変形例のサイト情報記憶部に記憶されている内容の一例を示す図である。

【図48】別の変形例のサイト情報記憶部に記憶されて

いる内容の一例を示す図である。

【図49】別の変形例の検索参照履歴記憶部に記憶されている内容の一例を示す図である。

【図50】別の変形例の表示部の一覧表示領域に表示される表示例を示す図である。

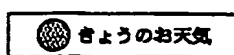
【図51】別の変形例の表示部の一覧表示領域に表示される別の表示例を示す図である。

【図52】従来の情報検索方法を説明するフローチャートである。

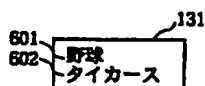
【符号の説明】

- 101 入力部
- 102 通信部
- 103 検索情報記憶部
- 104 関連語辞書
- 105 ハイパーリンク記憶部
- 106 検索部
- 107 表示情報作成部
- 108 学習部
- 109 表示部
- 121 通信検索部
- 122 通信バッファ
- 131 分野キーワード記憶部
- 132 トピックキーワード記憶部
- 133 検索起点記憶部
- 134 学習検索起点記憶部
- 135 検索履歴記憶部
- 161 検索制御部
- 162 検索初期化部
- 163 HTMLテキスト取得部
- 164 HTMLテキスト評価部
- 165 ハイパーリンク抽出部
- 166 ハイパーリンク評価部
- 167 検索終了判定部
- 168 ハイパーリンク選択部
- 181 学習制御部
- 182 学習バッファ
- 183 キーワード学習部
- 184 検索起点学習部
- 185 検索履歴学習部
- 4301 検索情報管理部
- 4302 検索情報制御部
- 4601 一括取得部
- 4602 Webページ記憶部

【図3】



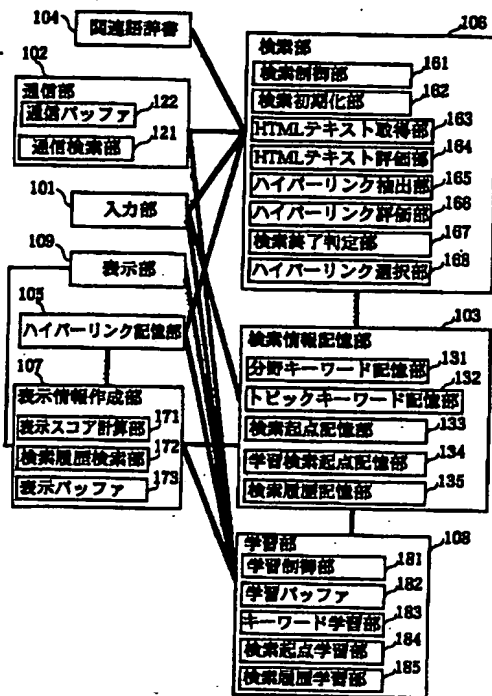
【図6】



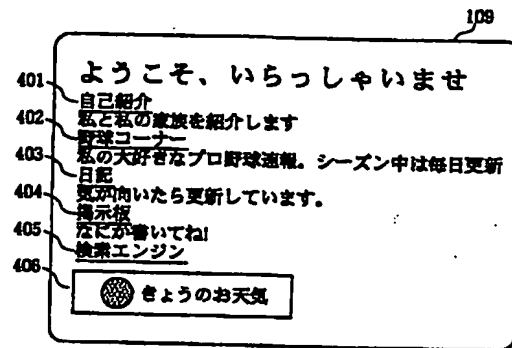
【図8】



【図1】



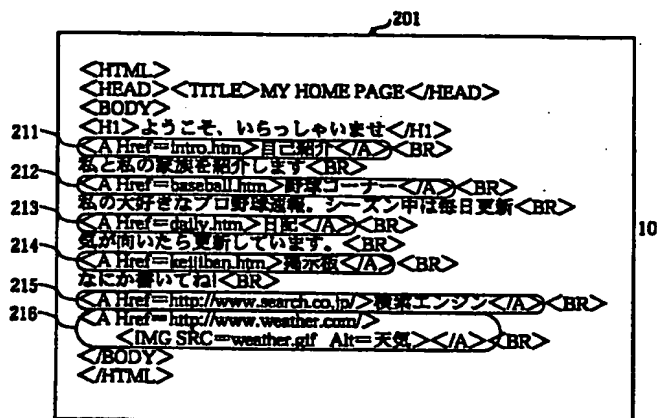
【図4】



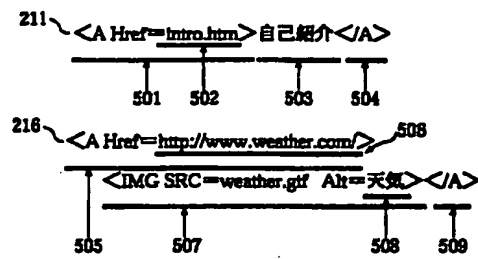
【図7】

701	半神	23	711
702	シャイアンツ	15	712
703	カーフ	12	713
704	ヘイスターズ	17	714
705	ヤグルト	08	715
706	プロ	02	716
707	高校	13	717
708	ゼリー	15	718

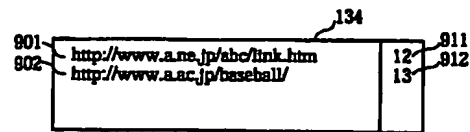
【図2】



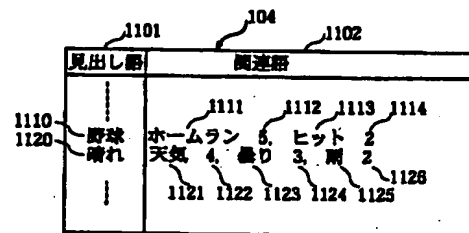
【図5】



【図9】



【図11】



【図10】

135

1001 http://www.a.ne.jp/abc/link.htm	2	1011	0	1021
1002 http://www.a.ac.jp/baseball/	3	1012	0	1022
1003 http://www.b.co.jp/base/tikera/	2	1013	1	1023
1004 http://www.c.co.jp/baseball/	7	1014	2	1024
1005 http://www.d.co.jp/pro/index.htm	2	1015	0	1025
1006 http://www.e.ne.jp/xyz/baseball/	1	1016	0	1026
1007 http://www.f.ac.jp/ca/hijk/inn/	1	1017	1	1027

【図12】

1200

	1201 URL	1202 取得フラグ	1203 リンク集フラグ	1204 抽出元番号	1205 アンカー文字列	1206 説明文字列	1207 ALT属性文字列	1208 ページスコア	1209 リンクスコア	1210 表示スコア
0	http://www.a.ne.jp/abc/	T	F	-1						
1										
2										
...										

【図13】

3000

	3001 URL	3002 アンカー文字列	3003 説明文字列	3004 表示スコア	3005 表示回数
0	http://www.a.ne.jp/abc/baseball.htm	野球コーナー		10	
1	http://www.b.co.jp/base/tikera/	タイカース堂	半導体情報誌のホームページ。個人成績が充実。	511	
2	http://www.c.co.jp/baseball/	HAPPY BASEBALL	ホームランにこだわったページ	501	
3	http://www.d.co.jp/pro/index.htm	プロ野球リンク	球団公式ページのリンク集	510	

【図14】

	1201	1202	1203	1204	1205	1400	1206	1207	1208	1209	1210
	URL	取得 フラグ	リンク 集フラグ	抽出元 番号	アンカー 文字列	説明 文字列	ALT属性 文字列	ページ スコア	リンク スコア	表示 スコア	
0	http://www.a.ne.jp/abc/	D	F	-1				.05			
1											
2											
3											

【図15】

	1201	1202	1203	1204	1205	1500	1206	1207	1208	1209	1210
	URL	取得 フラグ	リンク 集フラグ	抽出元 番号	アンカー 文字列	説明 文字列	ALT属性 文字列	ページ スコア	リンク スコア	表示 スコア	
0	http://www.a.ne.jp/abc/	D	F	-1				.05			
1	http://www.a.ne.jp/abc/ intro.htm	F		0	自己紹介						
2	http://www.a.ne.jp/abc/ baseball.htm	F		0	野球 コーナー						
3	http://www.a.ne.jp/abc/ daily.htm	F		0	日記						
4	http://www.a.ne.jp/abc/ keijiban.htm	F		0	掲示板						
5	http://www.search.co.jp/	F		0	検索エ ンジン						
6	http://www.weather.com/	F		0			天気				

【図16】

	1201	1202	1203	1204	1205	1500	1206	1207	1208	1209	1210
	URL	取得 フラグ	リンク 集フラグ	抽出元 番号	アンカー 文字列	説明 文字列	ALT属性 文字列	ページ スコア	リンク スコア	表示 スコア	
0	http://www.a.ne.jp/abc/	D	F	-1				.05			
1	http://www.a.ne.jp/abc/ intro.htm	F	F	0	自己紹介				0		
2	http://www.a.ne.jp/abc/ baseball.htm	F	F	0	野球 コーナー				10		
3	http://www.a.ne.jp/abc/ daily.htm	F	F	0	日記				0		
4	http://www.a.ne.jp/abc/ keijiban.htm	F	F	0	掲示板				0		
5	http://www.search.co.jp/	F	F	0	検索エ ンジン				0		
6	http://www.weather.com/	F	F	0			天気		0		

【図17】

	1201	1202	1203	1204	1205	1700	1206	1207	1208	1209	1210
	URL	取得 フラグ	リンク 集フラグ	抽出元 番号	アンカー 文字列	説明 文字列	ALT属性 文字列	ページ スコア	リンク スコア	表示 スコア	
0	http://www.sns.jp/abc/	D	F	-1				.05			
1	http://www.sns.jp/abc/ intro.htm	F	F	0	自己紹介				0		
2	http://www.sns.jp/abc/ baseball.htm	T	F	0	野球 コーナー				10		
3	http://www.sns.jp/abc/ daily.htm	F	F	0	日記				0		
4	http://www.sns.jp/abc/ keijiban.htm	F	F	0	掲示板				0		
5	http://www.search.co.jp/	F	F	0	検索エ ンジン				0		
6	http://www.weather.com/	F	F	0		天気			0		

【図18】

	1201	1202	1203	1204	1205	1800	1206	1207	1208	1209	1210
	URL	取得 フラグ	リンク 集フラグ	抽出元 番号	アンカー 文字列	説明 文字列	ALT属性 文字列	ページ スコア	リンク スコア	表示 スコア	
0	http://www.sns.jp/abc/	D	F	-1				.05			
1	http://www.sns.jp/abc/ intro.htm	F	F	0	自己紹介				0		
2	http://www.sns.jp/abc/ baseball.htm	D	F	0	野球 コーナー				10		
3	http://www.sns.jp/abc/ daily.htm	F	F	0	日記				0		
4	http://www.sns.jp/abc/ keijiban.htm	F	F	0	掲示板				0		
5	http://www.search.co.jp/	F	F	0	検索エ ンジン				0		
6	http://www.weather.com/	F	F	0		天気			0		

【図19】

1901

```

<HTML>
<HEAD><TITLE>Baseball</HEAD>
<BODY>
<h1>野球コーナー</h1>

<H2>昨日の試合結果</H2>

タイカース 3-2シャイアンツ<BR>
カーフ 4-1ヤグルト<BR>
ヘイスターズ3-2トラコンズ<BR>

<a href="http://www.sns.jp/abc/">戻る</a><BR>
<a href="/link.htm">リンク集</a><BR>
</BODY>
</HTML>
  
```

1903
1802

【図34】

	3401	3402	3403
	ダブル クリック回数	○評価ボタン クリック回数	×評価ボタン クリック回数
0	0	0	0
1	0	0	0
2	0	0	0
3	0	0	0

【図20】

	1201	1202	1203	1204	1205	2000	1206	1207	1208	1209	1210
	URL	取得 フラグ	リンク 集フラグ	抽出元 番号	アンカー 文字列	説明 文字列	ALT属性 文字列	ページ スコア	リンク スコア	表示 スコア	
0	http://www.s.ne.jp/abc/	D	F	-1				.05			
1	http://www.s.ne.jp/abc/ intro.htm	F	F	0	自己紹介				0		
2	http://www.s.ne.jp/abc/ baseball.htm	D	F	0	野球 コーナー			.33	10		
3	http://www.s.ne.jp/abc/ daily.htm	F	F	0	日記				0		
4	http://www.s.ne.jp/abc/ kazjiban.htm	F	F	0	掲示板				0		
5	http://www.search.co.jp/	F	F	0	検索エ ンジン				0		
6	http://www.weather.com/	F	F	0		天気			0		

【図37】

701	半神	29	711
702	シャイアンツ	15	712
703	カーフ	12	713
704	ヘイスターズ	17	714
707	高校	13	717
708	ゼリーグ	15	718
	情報	06	
	掲載	06	
	ホームページ	06	
	個人	06	

【図38】

【図21】

	1201	1202	1203	1204	1205	2100	1206	1207	1208	1209	1210
	URL	取得 フラグ	リンク 集フラグ	抽出元 番号	アンカー 文字列	説明 文字列	ALT属性 文字列	ページ スコア	リンク スコア	表示 スコア	
0	http://www.s.ne.jp/abc/	D	F	-1				.05			
1	http://www.s.ne.jp/abc/ intro.htm	F	F	0	自己紹介				0		
2	http://www.s.ne.jp/abc/ baseball.htm	D	F	0	野球 コーナー			.33	10		
3	http://www.s.ne.jp/abc/ daily.htm	F	F	0	日記				0		
4	http://www.s.ne.jp/abc/ kazjiban.htm	F	F	0	掲示板				0		
5	http://www.search.co.jp/	F	F	0	検索エ ンジン				0		
6	http://www.weather.com/	F	F	0		天気			0		
7	http://www.s.ne.jp/abc/ link.htm	F		2	リンク集						

701	半神	29	711
702	シャイアンツ	14	712
703	カーフ	11	713
704	ヘイスターズ	18	714
707	高校	12	717
708	ゼリーグ	14	718
	情報	06	
	掲載	06	
	ホームページ	06	
	個人	06	

【図30】

	3001	3002	3000	3003	3004	3005
	URL	アンカー文字列	説明文字列	表示 スコア	表示 回数	
0	http://www.s.ne.jp/abc/ baseball.htm	野球コーナー		10		
1	http://www.h.co.jp/base/ click/	タイカース堂	半神情報掲載のホームページ。 個人版が未定。	511		
2	http://www.c.co.jp/baseball/	HAPPY BASEBALL	ホームランにこだわったページ	501		
3	http://www.d.co.jp/pro/ index.htm	プロ野球リンク	球団公式ページのリンク集	510		

【図22】

	1201	1202	1203	1204	1205	2200	1206	1207	1208	1209	1210
	URL	取得 フラグ	リンク 集フラグ	抽出元 番号	アンカー 文字列	説明 文字列	ALT属性 文字列	ページ スコア	リンク スコア	表示 スコア	
0	http://www.a.ne.jp/abc/	D	F	-1				.05			
1	http://www.a.ne.jp/abc/ intro.htm	F	F	0	自己紹介				0		
2	http://www.a.ne.jp/abc/ baseball.htm	D	F	0	野球 コーナー			.33	10		
3	http://www.a.ne.jp/abc/ daily.htm	F	F	0	日記				0		
4	http://www.a.ne.jp/abc/ keijiban.htm	F	F	0	掲示板				0		
5	http://www.search.co.jp/	F	F	0	検索エ ンジン				0		
6	http://www.weather.com/	F	F	0		天気			0		
7	http://www.a.ne.jp/abc/ link.htm	F	T	2	リンク集				0		

【図23】

	1201	1202	1203	1204	1205	2300	1206	1207	1208	1209	1210
	URL	取得 フラグ	リンク 集フラグ	抽出元 番号	アンカー 文字列		説明 文字列	ALT属性 文字列	ページ スコア	リンク スコア	表示 スコア
0	http://www.a.ne.jp/abc/	D	F	-1					.05		
1	http://www.a.ne.jp/abc/ intro.htm	F	F	0	自己紹介					0	
2	http://www.a.ne.jp/abc/ baseball.htm	D	F	0	野球 コーナー				.33	10	
3	http://www.a.ne.jp/abc/ daily.htm	F	F	0	日記					0	
4	http://www.a.ne.jp/abc/ keijiban.htm	F	F	0	掲示板					0	
5	http://www.search.co.jp/	F	F	0	検索エ ンジン					0	
6	http://www.weather.com/	F	F	0			天気			0	
7	http://www.a.ne.jp/abc/ link.htm	T	T	2	リンク集					0	

【図31】

	3001	3002	3100	3003	3004	3005
	URL	アンカー文字列	説明文字列	表示 スコア	表示 回数	
0	http://www.a.ne.jp/abc/ baseball.htm	野球コーナー		10	0	
1	http://www.b.co.jp/base/ tickets/	タイカース党	半神情報掲載のホームページ。 個人成績が充実。	511	2	
2	http://www.c.co.jp/baseball/	HAPPY BASEBALL	ホームランにこだわったページ	501	7	
3	http://www.d.co.jp/pro/ index.htm	プロ野球リンク	球団公式ページのリンク集	510	2	

【図24】

	1201	1202	1203	1204	1205	2400	1206	1207	1208	1209	1210
	URL	取得 フラグ	リンク 集フラグ	抽出元 番号	アンカー 文字列	識別 文字列	ALT属性 文字列	ページ スコア	リンク スコア	表示 スコア	
0	http://www.a.ne.jp/abc/	D	F	-1				.05			
1	http://www.a.ne.jp/abc/ intro.htm	F	F	0	自己紹介				0		
2	http://www.a.ne.jp/abc/ baseball.htm	D	F	0	野球 コーナー			.33	10		
3	http://www.a.ne.jp/abc/ daily.htm	F	F	0	日記				0		
4	http://www.a.ne.jp/abc/ keijiban.htm	F	F	0	掲示板				0		
5	http://www.search.co.jp/	F	F	0	検索エ ンジン				0		
6	http://www.weather.com/	F	F	0			天気		0		
7	http://www.a.ne.jp/abc/ link.htm	D	T	2	リンク集				0		

【図25】

2501

```

<HTML>
<HEAD><TITLE>LINK</HEAD>
<BODY>
2501 <h1>野球リンク集</h1>
2502 <a href="http://www.b.co.jp/base/tickers/">タイガース</a><BR>
      半神情報満載のホームページ。個人成績が充実。<BR>
2503 <a href="http://www.c.co.jp/baseball/">HAPPY BASEBALL</a><BR>
      ホームランにこだわったページ<BR>
2504 <a href="http://www.d.co.jp/pro/index.htm">プロ野球リンク</a><BR>
      球団公式ページのリンク集<BR>
2505 <a href="http://www.e.ne.jp/xyz/baseball/">ハリーク王</a><BR>
      ハリークの情報が満載<BR>
2506 <a href="http://www.f.ac.jp/cs/hijk/inn/">三親王</a><BR>
      三振やエラーなど不名誉な記録を集めたページ<BR>
      <a href="http://www.a.ne.jp/abc/">戻る</a><BR>
</BODY>
</HTML>

```

【図39】

801	http://www.a.ne.jp/abc/link.htm	15	3901
802	http://www.a.ac.jp/baseball/	12	3902

【図40】

1001	http://www.a.ne.jp/abc/link.htm	2	1011	1	1021
1002	http://www.a.ne.jp/abc/link.htm	3	1012	1	1022
1003	http://www.a.ac.jp/baseball/	3	1013	0	1023
1004	http://www.b.co.jp/base/tickers/	8	1014	0	1024
1005	http://www.c.co.jp/baseball/	3	1015	0	1025
1006	http://www.d.co.jp/pro/index.htm	1	1016	1	1026
1007	http://www.e.ne.jp/xyz/baseball/	1	1017	2	1027

【図26】

	1201	1202	1203	1204	1205	2600	1206	1207	1208	1209	1210
	URL	取得 フラグ	リンク 集フラグ	抽出元 番号	アンカー 文字列	説明 文字列	ALT属性 文字列	ページ スコア	リンク スコア	表示 スコア	
0	http://www.a.ne.jp/abc/	D	F	-1				.05			
1	http://www.a.ne.jp/abc/ info.htm	F	F	0	自己紹介				0		
2	http://www.a.ne.jp/abc/ baseball.htm	D	F	0	野球 コーナー			.33	10		
3	http://www.a.ne.jp/abc/ daily.htm	F	F	0	日記				0		
4	http://www.a.ne.jp/abc/ halloween.htm	F	F	0	Halloween				0		
5	http://www.search.co.jp/	F	F	0	検索エ ンジン				0		
6	http://www.weather.com/	F	F	0		天気			0		
7	http://www.a.ne.jp/abc/ link.htm	D	T	2	リンク集			.11	0		

【図27】

	1201	1202	1203	1204	1205	2700	1206	1207	1208	1209	1210
	URL	取得 フラグ	リンク 集フラグ	抽出元 番号	アンカー 文字列	説明 文字列	ALT属性 文字列	ページ スコア	リンク スコア	表示 スコア	
0	http://www.a.ne.jp/abc/	D	F	-1				.05			
1	http://www.a.ne.jp/abc/ info.htm	F	F	0	自己紹介				0		
2	http://www.a.ne.jp/abc/ baseball.htm	D	F	0	野球 コーナー			.33	10		
3	http://www.a.ne.jp/abc/ daily.htm	F	F	0	日記				0		
4	http://www.a.ne.jp/abc/ halloween.htm	F	F	0	Halloween				0		
5	http://www.search.co.jp/	F	F	0	検索エ ンジン				0		
6	http://www.weather.com/	F	F	0		天気			0		
7	http://www.a.ne.jp/abc/ link.htm	D	T	2	リンク集			.11	0		
8	http://www.b.co.jp/ baseball.htm	F		7	野球 コーナー	野球コーナーのホーム ページ。個人成績が掲載					
9	http://www.a.co.jp/ baseball/	F		7	野球 コーナー	野球コーナーに こだわったページ					
10	http://www.d.co.jp/pro/ baseball.htm	F		7	野球 コーナー	野球コーナーの リンク集					
11	http://www.a.ne.jp/pro/ baseball/	F		7	野球 コーナー	野球コーナーの リンク集					
12	http://www.f.co.jp/abc/ halloween/	F		7	Halloween	Halloween ページなど 不審な記述を 含むページ					

【図47】

4701	http://www.a.ne.jp/	1201
4702	http://www.b.co.jp/	504
	http://www.c.co.jp/	702
	http://www.d.co.jp/	808

【図48】

4801	http://www.a.ne.jp/	20
4802	http://www.b.co.jp/	504
	http://www.c.co.jp/	72
	http://www.d.co.jp/	88

【図28】

	1201	1202	1203	1204	1205	2900 1206	1207	1208	1209	1210
URL	検索 フラグ	リンク 集フラグ	抽出 日時	アンカー 文字列	説明 文字列	ALT属性 文字列	ページ スコア	リンク スコア	表示 スコア	
0 http://www.s.na.jp/bbc/	D	F	-1				.05			
1 http://www.s.na.jp/bbc/bbcw.htm	F	F	0	自己紹介				0		
2 http://www.s.na.jp/bbc/bbcw.html	D	F	0	野球 コーナー			.33	10		
3 http://www.s.na.jp/bbc/daily.htm	F	F	0	日記				0		
4 http://www.s.na.jp/bbc/bc/index.htm	F	F	0	掲示板				0		
5 http://www.search.co.jp/	F	F	0	検索 エンジン				0		
6 http://www.weather.com/	F	F	0			天気		0		
7 http://www.s.na.jp/bbc/link.htm	D	T	2	リンク集			.11	0		
8 http://www.b.co.jp/	F	F	7	タイ ス	タイ特産品のホーム ページ、個人情報が不足			11		
9 http://www.s.co.jp/	F	F	7	HAPPY BASEBALL	ホームランに こだわったページ			1		
10 http://www.d.co.jp/pro/index.htm	F	T	7	プロ野球 リンク集	球団公式ページの リンク集			10		
11 http://www.s.na.jp/bbc/bbcw.html	F	F	7	リーグ	ハリークの情報が 豊富			0		
12 http://www.f.co.jp/bj/bj.html	F	F	7	三原王	三原王エラーなど 不十分な記録を 集めたページ			0		

【図29】

	1201	1202	1203	1204	1205	2900 1206	1207	1208	1209	1210
URL	検索 フラグ	リンク 集フラグ	抽出 日時	アンカー 文字列	説明 文字列	ALT属性 文字列	ページ スコア	リンク スコア	表示 スコア	
0 http://www.s.na.jp/bbc/	D	F	-1				.05		0	
1 http://www.s.na.jp/bbc/bbcw.htm	F	F	0	自己紹介				0	0	
2 http://www.s.na.jp/bbc/bbcw.html	D	F	0	野球 コーナー			.33	10	10	
3 http://www.s.na.jp/bbc/daily.htm	F	F	0	日記				0	0	
4 http://www.s.na.jp/bbc/bc/index.htm	F	F	0	掲示板				0	0	
5 http://www.search.co.jp/	F	F	0	検索 エンジン				0	0	
6 http://www.weather.com/	F	F	0			天気		0	0	
7 http://www.s.na.jp/bbc/link.htm	D	T	2	リンク集			.11	0	0	
8 http://www.b.co.jp/	F	F	7	タイ ス	タイ特産品のホーム ページ、個人情報が不足			11	511	
9 http://www.s.co.jp/	F	F	7	HAPPY BASEBALL	ホームランに こだわったページ			1	501	
10 http://www.d.co.jp/pro/index.htm	F	T	7	プロ野球 リンク集	球団公式ページの リンク集			10	510	
11 http://www.s.na.jp/bbc/bbcw.html	F	F	7	リーグ	ハリークの情報が 豊富			0	0	
12 http://www.f.co.jp/bj/bj.html	F	F	7	三原王	三原王エラーなど 不十分な記録を 集めたページ			0	0	

【図50】

【コンテンツ】

- 近畿の天気
- 大阪名物たこやき屋
- 大阪定食B
- 大阪定食A

【目次】

- お好み焼き「新大阪」

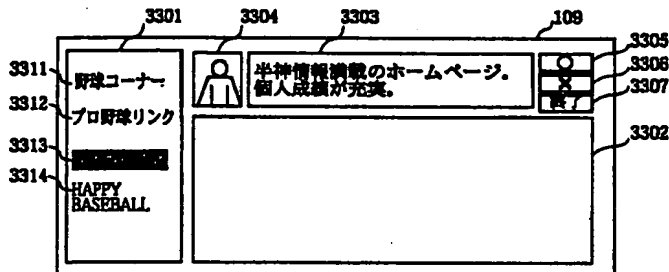
【リンク集】

- 大阪の食事処情報

【図32】

	3001 URL	3002 アンカー文字列	3200 説明文文字列	3003 表示 スコア	3004 表示 回数
0	http://www.a.ne.jp/abc/ baseball.htm	野球コーナー		10	0
1	http://www.d.co.jp/pro/ index.htm	プロ野球リンク	球団公式ページのリンク集	510	2
2	http://www.b.co.jp/base/ tikara/	タイカース党	半神情報満載のホームページ。 個人成績が充実。	511	2
3	http://www.c.co.jp/baseball/	HAPPY BASEBALL	ホームランにこだわったページ	501	7

【図33】



【図35】

	3401 ダブル クリック回数	3402 ○評価ボタン クリック回数	3403 ×評価ボタン クリック回数
0	1	0	0
1	1	0	1
2	1	1	0
3	0	0	0

【図36】

(a)	0 野球、コーナー
	1 プロ、野球、リンク、球団、公式、ページ
	2 タイカース、半神、情報、満載、ホームページ、個人、成績、充実
	3 ホームラン、ページ

(b)
タイカース、半神、情報、満載、ホームページ、個人、成績、充実
野球、コーナー、プロ、リンク、球団、公式、ページ

(c)
タイカース、半神、情報、満載、ホームページ、個人、成績、充実

(d)
半神、情報、満載、ホームページ、個人、成績、充実

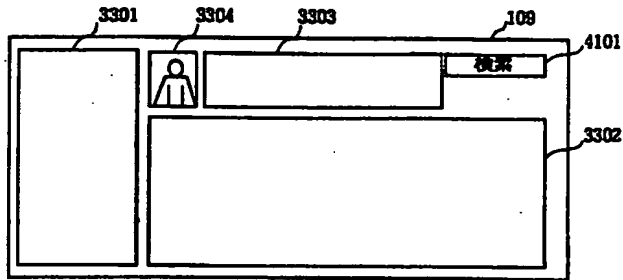
(e)
半神 6、情報 6、満載 6、ホームページ 6、個人 6、成績 6、充実 6

(f)
半神 6、情報 6、満載 6、ホームページ 6、個人 6

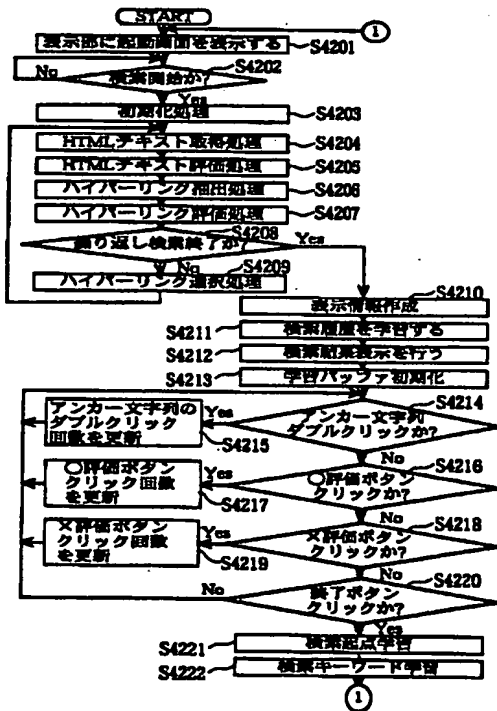
【図49】

4901	4902
http://www.a.co.jp/sub4.htm	3
http://www.a.co.jp/osaka.htm	1
http://www.b.co.jp/	0
http://www.c.co.jp/	0
http://www.d.co.jp/	2
http://www.c.co.jp/menu1.htm	1
http://www.c.co.jp/menu2.htm	2
http://www.a.co.jp/sub1.htm	5
http://www.a.co.jp/sub2.htm	6
http://www.a.co.jp/sub3.htm	7
http://www.a.co.jp/sub5.htm	2
http://www.a.co.jp/sub6.htm	0
http://www.a.co.jp/	1
http://www.a.co.jp/kobe.htm	0
http://www.a.co.jp/1ja.htm	2

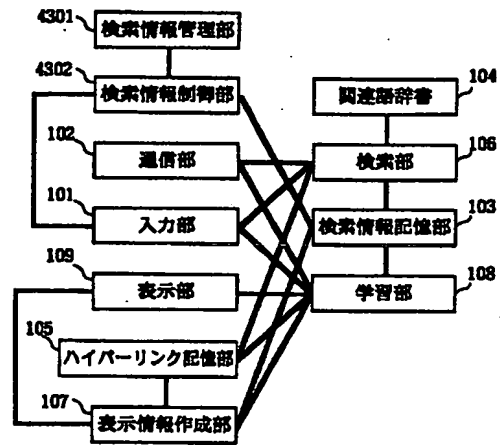
【図4 1】



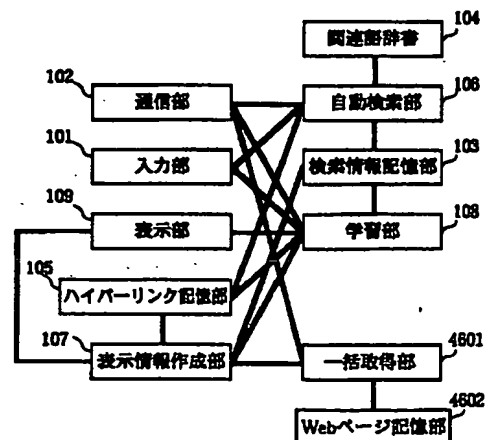
【図4 2】



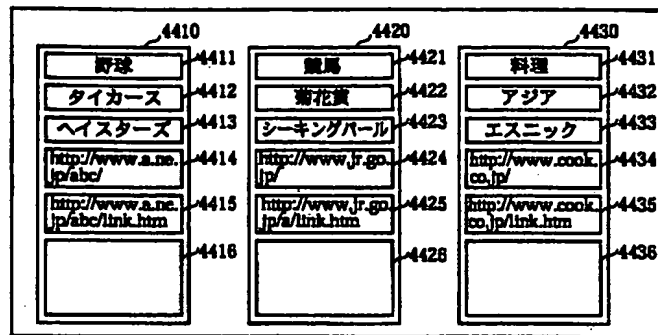
【図4 3】



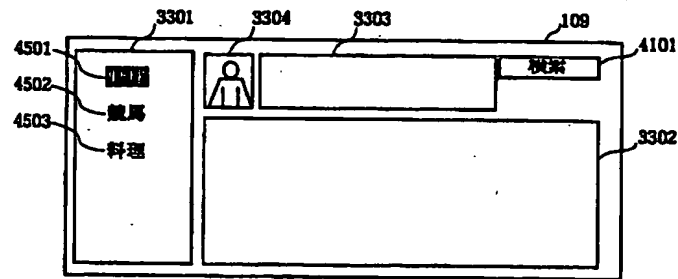
【図4 6】



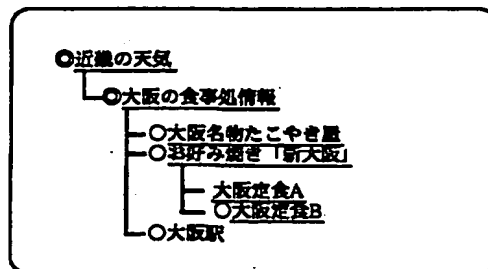
【図44】



【図45】



【図51】



【図52】

